THE FAULT IS IN OUR TCRS: SEARCHING FOR ON AND OFF-TARGETS IN T CELL RECEPTOR BASED THERAPIES.

A Dissertation

Presented to the Faculty of the Weill Cornell Graduate School

of Medical Sciences

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Ron S. Gejman

August 2018

THE FAULT IS IN OUR TCRS: SEARCHING FOR ON AND OFF-TARGETS IN T CELL RECEPTOR BASED THERAPIES

Ron S. Gejman, PhD

Cornell University 2018

T cell receptor (TCR)-based therapeutic cells and agents have emerged as a new class of effective cancer therapeutics against intracellular cancer-associated proteins. These agents rely on presentation of short peptides derived from cellular, viral or phagocytosed proteins on major histocompatibility complex (MHC). However, cross-reactivities of these agents to off-target cells and tissues are poorly understood, difficult to predict, and have resulted in serious, sometimes fatal, adverse events. We have developed a mammalian, minigene-based method (termed “PresentER”) that encodes MHC-I peptide ligands for functional immunological assays as well as for determining the reactivities and potential cross-reactivities of TCR-like therapeutic agents against libraries of MHC-I ligands. This system is highly specific to, and entirely dependent on, the genetically encoded MHC peptide sequence, because it does not require proteasome cleavage, transporter associated with antigen processing (TAP) or processing, for immune presentation. Cells expressing PresentER antigens can be bound by TCR and TCR mimic (TCRm) antibodies, activate antigen-specific T cells, lead to antigen-specific cell death *in vitro* and tumor rejection *in vivo*. Using PresentER in a pooled library screen, we find dozens of MHC-I ligands encoded in the human proteome that are cross-reactive with two TCR mimic antibodies and are not predictable by other methods. We extend the use of this method to find the targets of an engineered TCR.

Finally, we leverage the ability to generate tumors comprised of libraries of MHC-I ligands to study the determinants of MHC-I peptide immunogenicity *in vivo*. Surprisingly, we show that highly immunogenic tumor antigens (encoded by PresentER) do not lead to early immune-mediated tumor rejection *in vivo* when the fraction of cells bearing each antigen ("clonal fraction") is low. Moreover, the clonal fraction that must bear an antigen in order to lead to rejection of immunogenic tumor subclones is dependent on the individual antigen itself. These data indicate that tumor neoantigen heterogeneity has an underappreciated impact on the ability of the immune system to detect and eliminate cancer cells and has implications for the prioritization of antigens in the design of novel antigen-specific immunotherapeutics such as cancer vaccines.

BIOGRAPHICAL SKETCH

Ron was born in Rockville, Maryland to Argentine parents: a physician-scientist father and a Jewish educator mother. His father inculcated in Ron an early interest in logic and philosophy and later science and medicine. Beginning at the age of 11-12, Ron taught himself how to program computers, which opened an avenue by which to do scientific research while in high school. His father helped him secure a mentored summer position with Dr. Maria Martinez working on tools to analyze genome-wide association studies. He continued onto Columbia University for his Bachelor of Arts, starting in 2006 and double majoring in Philosophy and Biology. During college, he became very interested in RNA, a molecule which is at the very center of molecular biology, and worked in the laboratory of Dr. Lawrence Chasin for 3 years studying the regulation of RNA splicing using publically available DNA and RNA sequencing data. The explosion of data from the human genome project and other large data-collection projects meant that there were many questions that could now be asked and answered by just querying datasets online, without having to lift a pipette, but with the caveat that data could be wrong or biased in unanticipated ways. Of all of that Ron learned from Dr. Chasin's lab, the most critical were how to design simple experiments to test an incisive question and how to consider and deal with confounders when analyzing large data. In 2010, after graduating from college, he worked for one year in the laboratory of Dr. Alexander Tarakhovsky studying chromatin biology using ChIP-Seq and running the lab's Illumina HiSeq. Ron became enthralled with the immune system as a model system within which to learn about all major principles in biology, from molecular concepts to the organization of large numbers of cells into tissue. With access to a HiSeq pumping out massive amounts of DNA sequences every week, he realized that his training in computational biology

made him uniquely well suited to do “high throughput” biology for discovery and therapeutic purposes.

Ron decided to pursue an MD/PhD at the Weill Cornell Medical College / Rockefeller University / Memorial Sloan Kettering Cancer Center Tri-Institutional MD/PhD program, matriculating in 2011. He began his PhD in Dr. David Scheinberg’s lab and became fascinated with peptide antigen presentation and immunogenicity, trying to understand why certain mutations made cancer cells recognizable to the immune system and other did not. The combinatorial problem posed by immune discrimination between altered/foreign and self-peptides was the subject of Ron’s interests for the next 5 years. He became excited by the possibility of harnessing the immune system’s tools to selectively kill cancer cells on the basis of altered peptide ligands they present on the surface of their cells was, but he recognized that it would be difficult to ensure these drugs were safe before they were given to a person. Therefore, he developed a method to express libraries of peptide-MHC molecules in mammalian cells and applied this approach to systematically identify ligands of cancer therapeutics that might make them unsafe to give to humans. He then turned to the study of immunogenicity *in vivo* models of cancer to try and ask and answer his original question: what are the characteristics of immunogenic MHC-I neoantigens?

To my parents, Alicia and Pablo, who "watered me" and to my sister, Maya, who tried to keep me grounded.

To Shira—my wife and partner—who reminds me that life is also a playground.

To my scientific mentors Larry Chasin and David Scheinberg—who have left their imprint on my thinking.

ACKNOWLEDGEMENTS

We thank Eureka Therapeutics for their generous gift of the ESK1 and Pr20 antibodies; Brian Baker and Lance Hellman for the gift of A6 TCR plasmids and Yael David for assistance in expressing and refolding; Scott Lowe and Eusebio Manchado Robles for the gift of the MSCV backbone and assistance with pooled library cloning; Steven A. Rosenberg for the DMF5 TCR. Thank you to Andrew Scott for advice and help with flow cytometry staining panels; Justin Mulvey for assistance with analysis of pathology slides. Many thanks to Taha Merghoub, Mathieu Gigoux, Martin Klatt, Hans Schreiber and Karen Schreiber for useful conversations We graciously acknowledge the help and support of the Integrated Genomics Operation (in particular Kety Huberman and Juan Li), Flow Cytometry and Molecular Cytology core facilities at Memorial Sloan Kettering.

Thank you to my final thesis committee members: David Scheinberg, Andrea Schietinger, Alex Kentsis, Gary Koretzky and Daniel Mucida.

Thanks to Aaron Chang, Heather Jones, Harlan Pietz, Tatiana Korontsvit, Claire Oh, Victoria Zakhaleva, Emily Casey, Casey Jarvis, Nicholas Veomett, Melissa Mathias, Thomas Gardner, Michael McDevitt and Tao Dao for technical, experimental and theoretical assistance.

The projects described in this dissertation were supported by the Parker Institute for Cancer Immunotherapy (Chapter I) and the Functional Genomics Initiative at Memorial Sloan Kettering Cancer Center (Chapter II). Ron S. Gejman was directly supported by NCI F30 CA200327 and NIGMS T32GM07739.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

DMSO	Dimethyl sulfoxide
ER	Endoplasmic reticulum
ESK1	anti-WT1 126-134/HLA-A2
HAM/TSP	HTLV-1-associated myelopathy/tropical spastic paraparesis
HLA	Human leukocyte antigen
HTLV-1	Human T-lymphotropic virus 1
mAb	Monoclonal antibody
MHC	Major histocompatibility complex
Ova	Chicken ovalbumin protein or peptide
PBMC	Peripheral Blood mononuclear cells
pMHC	MHC presented peptides; Peptide-MHC
Pr20	anti-PRAME300-309/HLA-A2 clone 20
PRAME	Preferentially expressed antigen in melanoma
TAP	Transporter associated with peptide processing
TAPBPR	Tapasin related protein
TCR	T cell receptor
TCRm	T cell receptor mimic antibody
TEIPP	T cell epitopes associated with impaired peptide processing
TIL	Tumor infiltrating lymphocyte
WT1	Wilms Tumor protein 1
β 2M	beta-2-microglobulin

INTRODUCTION

The basic premise of cancer therapy is to kill cancer cells while sparing normal cells. The simplicity of this premise is belied by the difficulty of finding cellular characteristics that both efficiently distinguish cancerous from normal cells and can be effectively targeted by therapeutic agents. While cancer cells differ from other cells in many respects, most of the differences are found within the intracellular space, and thus inaccessible to most potential therapeutics (from small molecules to antibodies to engineered cells). In response to the challenge of developing molecules that can cross the cellular plasma membrane, recent successful cancer therapeutics have focused on the small subset of proteins that are found on the surface of cancer cells[1]. However, the past 50 years have shed light on a generalized molecular mechanism found in all nucleated cells that exposes fragments of cellular proteins on the surface of the plasma membrane[2]. The discovery of this mechanism, now called antigen presentation, has ushered in a new paradigm in the treatment of cancer and will be the subject of this dissertation.

The Major histocompatibility complex (MHC) is a cell-surface protein that is assembled in the endoplasmic reticulum together with beta-2-microglobuin (β2M) and a short, variable protein fragment. In humans, MHC molecules are also referred to Human leukocyte antigen (HLA). The protein fragment (peptide) is noncovalently embedded in the groove of the MHC/β2M complex and the entire complex is exported to the surface of the cell. The peptides presented on MHC are derived from proteolytic cleavage of many cellular proteins, thus thousands of distinct peptide-MHC (pMHC) can be found the plasma membrane[3,4]. There are two flavors of MHC

molecules: MHC-I, which is found on all nucleated cells and generally presents 8-11 amino acid peptides and MHC-II, which is found primarily on specialized antigen presenting cells such as dendritic cells, macrophages and B cells. MHC-II molecules presents longer peptides up to about 20 amino acids long[5].

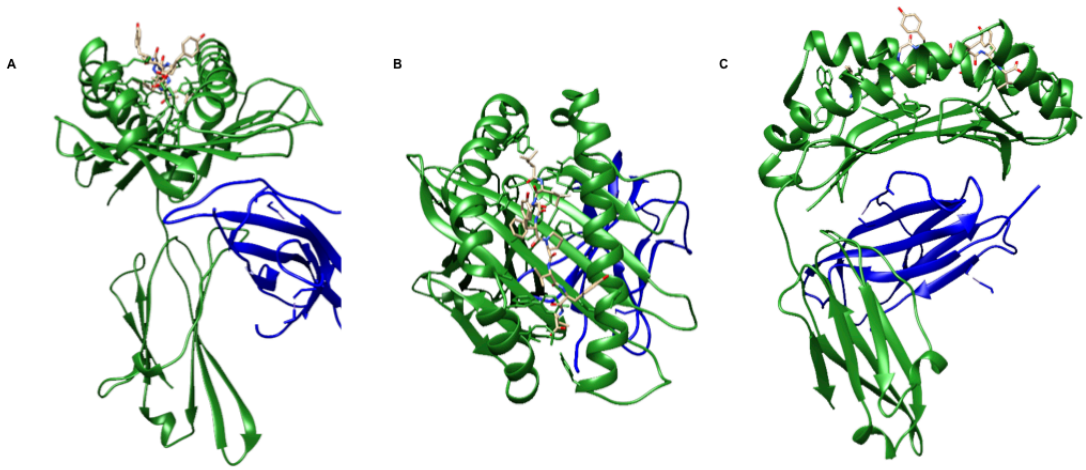


Figure 1 Structure of HLA-A*02:01 with peptide. The structure of HLA-A*02:01 in complex with beta-2-microglobulin and a peptide from the Human T-lymphotropic virus type I (HTLV-1) Tax (11-19) LLFGYPVYV. Structures downloaded from the RCSB Protein Data Bank[6] ID # 1AO77

There is tremendous genetic variability at the MHC locus in the human population, affecting the size, shape and biochemical characteristics of the presented peptides. In the human genome there are three classical MHC-I genes: HLA-A, HLA-B and HLA-C. Across all three of these genes there are >13,000 known alleles (as of July, 2018)[8]. As a diploid organism, humans can express as few as three and as many as six distinct classical MHC-I alleles. The complete repertoire of peptides that are presented on a human cell is based on an individual's genetic diversity at the MHC locus as well as the genes expressed by that cell. For instance, the most common MHC-I allele in the Caucasian population of the United States is HLA-A*02:01, which typically binds and presents peptides of 9 amino acids in length with hydrophobic

amino acids in the 2nd and 9th position **(Figure 2)**. Another common allele, HLA-A*24:02 binds preferentially to peptides with a Y/F/W/M in the 2nd position and a F/I/L in the 9th position. Thus, a person that is homozygous for HLA-A*02:01 will present different peptides from a person who is homozygous for HLA-A*24:02. It is not only the motif of peptides presented on HLA that differs by allele, but also the repertoire size, affinity diversity of peptides presented by each HLA allele[9].

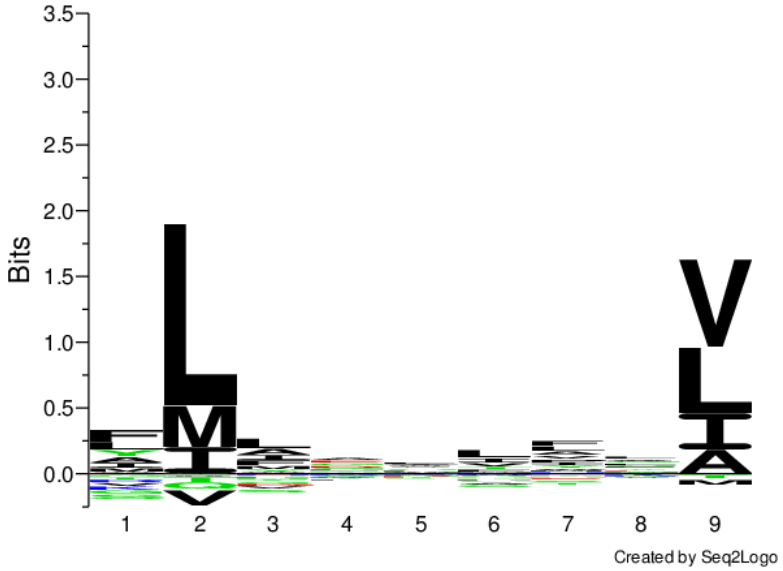


Figure 2 HLA-A*02:01 peptide binding motif. The MHC-I molecule HLA-A*02:01 typically binds linear peptides with hydrophobic amino acids in the 2nd and 9th position. This motif was generated using NetMHCPan v4.0 based on known HLA-A*02:01 ligands.

Cell state and health can dramatically change the repertoire of presented peptides. For instance, infection with a virus leads to presentation of peptides derived from viral proteins, in addition to the endogenous peptides that are normally presented[10]. A general principle is that the most abundant peptides found on surface MHC are those with high affinity to the MHC molecule and encoded by highly transcribed and translated genes[11]. Indeed, low affinity peptides can be replaced with higher affinity

peptide using TABPR, which displaces low affinity peptides and stabilizes empty MHC molecules until a higher affinity ligand is found[12].

MHC presented peptides (pMHC) are announcements to the immune system about which proteins are expressed within the cell. For instance, cells infected with intracellular pathogens (viruses, mycobacteria) will express foreign proteins and their proteolytic degradation may yield MHC ligands. T cells that recognize the foreign peptides displayed on the cell can activate and kill the infected cell. The general principle by which T cells recognize target cells is through engagement of the T cell receptor (TCR) on the pMHC. Each T cell develops a distinct, semi-randomly generated TCR while they mature in the thymus. T cells bearing non-functional TCRs or TCRs that bind to self-peptides presented on MHC are clonally deleted while T cells with functional, non-auto-reactive TCRs are permitted to leave the thymus. T cells that encounter a cognate pMHC antigen can, given the appropriate cytokines and contextual cues, activate, proliferate and kill cells bearing the foreign pMHC.

The diversity of T cells and the genetic heterogeneity of peptide-MHC is thought to reflect an evolutionary arms race between pathogens and organisms with an adaptive immune system. In brief, populations of pathogens are continuously infecting populations of mammals and evolving to escape from the immune system. In turn, mammals have developed high levels of genetic heterogeneity at the MHC locus in order to prevent the emergence of pathogens with protein sequences that cannot be bound by any MHC allele in the population. Individuals carrying multiple genetically dissimilar MHC-I alleles can present a broader range of all possible short peptides[13], and are thus theoretically more resilient to pathogens. In order to recognize and kill cells expressing foreign MHC-I peptides, mammals have evolved the ability to

generate an enormous and diverse T cell repertoire where each T cell can theoretically recognize a different pMHC. A structure of a TCR in complex with its target peptide presented on MHC-I is shown in **Figure 3.**

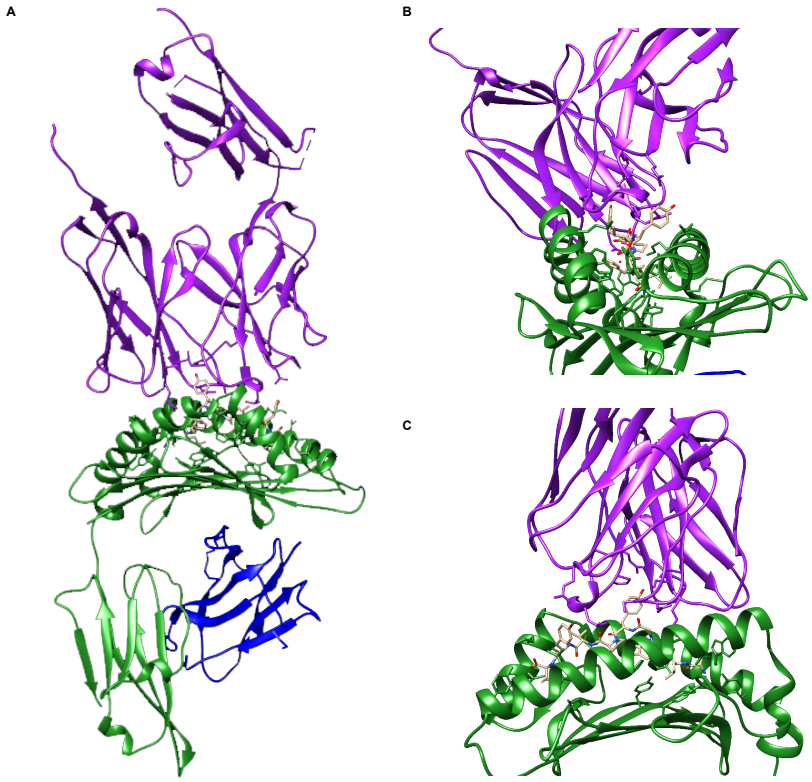


Figure 3 Structure of the TCR-peptide-HLA ternary complex. The structure of the A6 TCR bound to a peptide from the Human T-lymphotropic virus type I (HTLV-1) Tax (11-19) LLFGYPVYV presented on HLA-A02:01. Structure downloaded from the RCSB Protein Data Bank[6]; ID # 1AO7[7]

The diversity of distinct pMHC that can be recognized by T cells allows T cells to respond to nearly all pathogens that might enter the organism, including those that have never been seen before. However, there is an upper limit to the number of T cells found in any individual. Humans are estimated to have ~10^{12} T cells, but this number is several orders of magnitude lower than the number of possible MHC-I and

MHC-II ligands. In order to deal with the enormous diversity of possible target sequences, TCRs are cross-reactive[14-16]. A single TCR is estimated to be capable of binding to 1 million distinct pMHC[16]. The risk that such a degenerate system of pathogen recognition poses is that a T cell may bind to both a pathogenic peptide as well as a self-peptide, thus activating in response to a pathogen, but also causing autoimmune disease. In practice, widespread cross-reactivity does not lead to autoimmune disease because most autoreactive T cells are clonally eliminated during T cell development in the thymus[17,18]. Autoreactive T cells that escape the thymus can be rendered nonfunctional by several mechanisms; for instance, if they do not receive the appropriate contextual cues ("danger signals") at the same time as they encounter their cognate antigen.

T cells are not restricted to the recognition and killing of cells infected with pathogens. They can also recognize cancer cells when those cells display mutated peptides or overexpressed cancer-specific self-peptides. T cells found within tumors can recognize somatic mutation-derived neoantigens expressed and presented by tumor cells. T cells that recognize neoantigens have even been used to successfully treat patients. One approach that has been used is transfer of autologous tumor infiltrating lymphocytes (TILs) into patients. Briefly, this is done by harvesting tumor infiltrating lymphocytes (TILs), expanding them *ex vivo* and re-infusing them into patients[19-21]. While successful in some cases, TIL based therapies are patient-specific, costly and efficacy is difficult to evaluate because each patient receives a different "personalized" treatment. An alternative approach is to target "universal" cancer-specific pMHC that are found across many different patients and tumor types. Many therapies directed towards these universal pMHC targets have been developed, including T cells expressing engineered TCR[22-24] or antibodies that mimic the

structure of TCR (TCRm)[25-28]. Engineered TCR and TCR mimic antibodies are more cost effective and easier to test in clinical trials because they do not vary from patient-to-patient. For instance, Wilms Tumor 1 (WT1) is a protein that is widely expressed in cancers but only expressed in a minority of normal tissues. HLA-A*02:01^{+}/WT1^{+} cells express a specific WT1-derived pMHC on the surface: RMFPNAPYL. Several RMFPNAPYL/HLA-A*02:01 specific therapeutic drugs have been developed, including T cells expressing affinity matured TCRs[29] and antibodies that have been selected to bind to this pMHC (TCR mimic antibodies)[30].

Figure 4 presents a schematic of how MHC presented peptides are generated and interact with therapeutic agents such as T cells and TCR mimic antibodies.

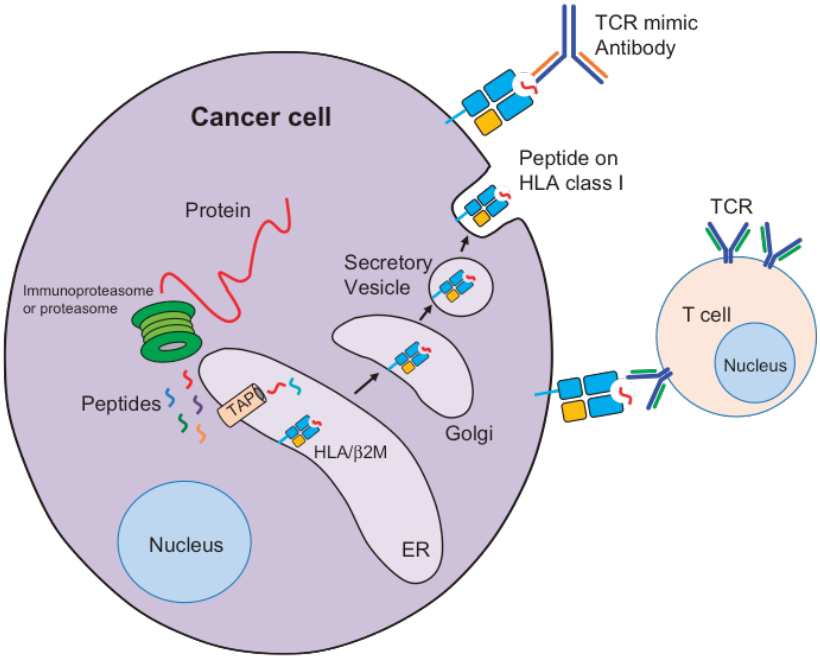


Figure 4 MHC-I antigen presentation and antigen recognition by T cells and TCR like molecules. Cancer cells can present cancer-specific peptides on the surface that can be bound to by T cells and TCR like molecules such as TCR mimic antibodies. Reprinted from Chang et al[25].

Whether an MHC-I molecule is presenting a viral peptide or a self-peptide, the vast majority of the binding surface presented to a T cell is identical. The only region of the MHC that is “specific” and thus marks a cell as virally infected or as a cancer is the linear peptide found in the MHC groove. The groove is the region of the complex to which TCRs bind (**Figure 3**) and TCR specificity is generated by engaging with the residues of the peptide that face “up” from the MHC groove. The same MHC-derived structural constraints apply to other biologic molecules that engage with MHC presented peptide. For instance, an antibody that binds to a specific peptide presented on MHC must engage primarily with the upward facing residues of the peptide. If it binds too strongly to the invariant portion of the MHC then it will necessarily be cross-reactive. Thus, although TCRs are low affinity (μM), cell-anchored and monovalent (one TCR binds to one epitope) whereas antibodies are soluble, high affinity (nM) and bivalent (one antibody binds to two identical targets), the mechanism by which they must recognize their target is very similar. From the perspective of the total binding surface, two different peptides presented on HLA-A*02:01 are highly similar. This is in contrast to traditional antibody targets, where epitopes from distinct proteins are typically dissimilar. Thus, making therapeutic antibodies against a traditional, extracellular or soluble target protein is a significantly less challenging than making a TCR like molecule that is specific to a single presented MHC-I peptide.

TCR-based therapeutics that do not undergo negative selection for the human pMHC repertoire or are further engineered for high affinity binding can be auto-reactive. A prominent example is an affinity-enhanced engineered T cell directed towards the MAGE-A3 peptide HLA-A*01:01/EVDPIGHLY, which induced lethal cardiotoxicity in

two patients during a phase I clinical trial. Although this TCR was originally located in a human, it was modified *in vitro* to increase its affinity for the MAGE-A3 peptide. Extensive preclinical testing failed to uncover off-target reactivity of the anti-MAGE TCR, but afterwards it was discovered that an epitope derived from the Titin protein (a structural protein highly expressed by cardiomyocytes), was cross-reactive with the MAGE-A3 TCR (HLA-A*01:01/ESDPIVAQY)[31]. Hence, a major challenge to the development of safe TCR based therapeutics is the prospective identification of off-tumor off-targets[25].

In mammals, an entire organ (the thymus) is used to eliminate auto-reactive T cells. However, for drug developers, there is no currently available mechanism to negatively select TCR like therapeutic molecules during drug development. We cannot yet screen a candidate TCR drug against a wide range of target epitopes to identify cross-reactivities. For traditional antibodies, this is acceptable because antibodies are not as likely to be cross reactive as TCR mimic antibodies and because antibodies can be tested for safety in animals (where protein homology means that at least some cross-reactivities will be detected) as well as in dose-escalation clinical trials in humans. However, TCR-like molecules, which are likely to be cross-reactive against unknown antigens presented on human-specific MHC molecules for the reasons specified above, cannot easily be tested in animal models because they will not bind to non-human MHC. Moreover, although TCR-like monoclonal antibodies can theoretically be dose-escalated to evaluate toxicity, TCR mimics of this format have not yet been used in clinical trial. The only TCR-like molecules that have gone into humans are native or affinity matured T cell receptors that have been heterologously expressed in donor-patient T cells and infused into patients, where they undergo

rapid clonal expansion and are thus not easily tested for safety by dose-escalation studies.

Identifying off-tumor, off-targets of TCR like therapeutics is challenging for three reasons: (1) the scope and extent of the repertoire of MHC ligands in humans is unknown, despite multiple reports of isolation by mass spectrometry, (2) cross-reactive pMHC are not readily predictable from crystal structures or alanine scanning and (3) animal models of cross-reactivity are not possible due to the species-specific peptide processing[32] and structure of MHC. Methods to search the MHC-I ligandome for TCR targets have been developed with yeast[33,34], insect-baculovirus[35] and tetramer[36,37] technologies. These systems are powerful because billions of MHC-I ligands can be encoded by a population and used to find cross-reactivities—and they have been successfully used to discover cross-reactive targets of some TCRs, in addition to elucidating fundamental biology of TCRs[38]. However, the cellular systems use a synthetic covalent linker to enforce peptide-MHC proximity, which allows presentation of peptides that would not ordinarily be presented and may distort the structure of the epitope. The tetramer-based screening systems rely on peptide synthesis, which is expensive and time consuming. Furthermore, tetramer, yeast and insect systems cannot be used for *in vitro* and *in vivo* immunology assays, such as those that measure T cell activation and cytotoxicity.

The first chapter of this dissertation deals with the development of a method, termed "PresentER," to express single peptide-MHC in mammalian cells and enable screening of TCR and TCR like molecules against thousands of possible off-target peptides. The second chapter of this dissertation establishes that PresentER encoded antigens can be used to study tumor immunology *in vivo* using

immunocompetent animal models to discover the determinants of peptide immunogenicity.

Chapter I: Development of an MHC-I peptide presentation system

Introduction

The number of endogenous MHC-I ligands that are theoretically presentable in human tissue is enormous. Using an *in silico* digest of the human proteome and the MHC-I ligand predictor NetMHCPan, over 750,000 peptides of 9 and 10 amino acids in length are predicted to bind to HLA-A*02:01 (ic50 < 500nM). Over the thousands of MHC-I alleles found in the human population, many of which have slight variations on the peptides they present, the number of peptides that can be presented increases dramatically. However, not all peptides that are predicted to bind to MHC-I *in silico* will be found on the surface of a cell. For instance, some peptides are cleaved in the middle by the proteasome or immunoproteasome[25], inefficiently transported by TAP[38], found in low abundance[10,39], or are never generated in the first place because they are not cleaved at the C-terminus[38]. Some peptides are generated by the proteasome and bind to MHC, but are replaced with higher affinity peptides in the ER by the "peptide editor" TAPBPR[11].

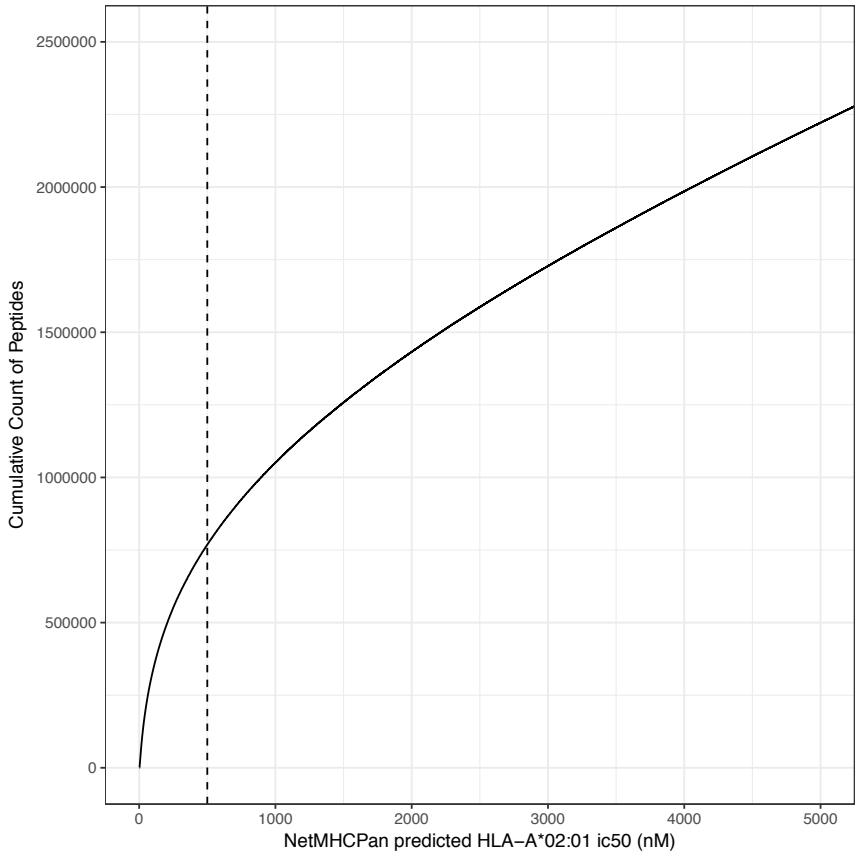


Figure 5 Cumulative number of predicted HLA-A*02:01 peptides across the human proteome. The complete human proteome downloaded from UniProt (May, 2017) was digested *in silico* into 9mer and 10mer peptide sequences. The predicted affinity to HLA-A*02:01 was calculated with NetMHCPan v4.0 and the cumulative number of peptides predicted to bind at each affinity is plotted. Affinities are represented as 50% inhibitory concentrations. Lower ic50s are higher affinity. A dashed line at 500 nM indicates the empirically determined cut-off for HLA-A*02:01 binding.

Major advances have been made in the past few years to identify the MHC-I ligands that are actually presented on human cells. The general approach taken is to immunoprecipitate HLA complexes from isolated plasma membranes using an antibody that binds HLA and disassociate the peptides from MHC by acid washes. The eluted peptides are chromatographically separated and subject to tandem mass spectrometry. This approach has led to the discovery of hundreds of thousands of

presented peptides across all major HLA types[4,10,40-42]. However, it remains unclear how many presented peptides remain to be discovered. In other words: the actual, complete ligandome of MHC-I peptides is unknown. Moreover, little is known about the antigens presented on important tissues where access is limited: brain, eye, heart, etc. These tissues may present peptides which can cross-react with TCR therapeutics. Another challenge is determining the ligandome of cells present in low abundance (e.g. retina). Thus, until there is more confidence that the complete repertoire of presented human peptides is known, it remains difficult and work-intensive to ensure that a TCR like therapeutic molecule will not cross-react with an endogenously expressed and presented pMHC. Indeed, large numbers of peptides that are derived from human proteins and predicted to bind to MHC-I must currently be tested to be relatively certain that a cross-reactive peptide is not found.

The classical method used to validate that a TCR like therapeutic molecule is specific to its target is to assay its binding to a small number of rationally selected off-targets. First, a small number of peptides are synthesized in order to determine where the TCR like molecule binds along the linear peptide target. Generally, each of these peptides differs from the target peptide by a single amino acid substitution. For instance, each amino acid can be replaced with an alanine (also known as an "alanine scan"). These peptides are resuspended in an amphipathic solvent such as DMSO and pulsed at molar excess onto TAP deficient cells (e.g. 500μM) together with soluble beta-2-microglobulin, where they displace the peptides found on the MHC and allow refolding with the new peptide to occur[44]. This is known as "peptide pulsing" or alternatively "MHC stabilization assay" because the peptide stabilizes MHC on the surface of the cell, leading to an increase in the total amount of surface MHC[44]. Based on the peptide variants that bind to the TCR like molecule or activate T

cells expressing the TCR like molecule, new sets of peptides are designed to probe the specific residues that are of functional importance to the binding. The cost to synthesize individual peptides at microgram scale is ~$7-10 per amino acid ($70-$100 per peptide). Therefore, testing even a few peptides quickly becomes cost prohibitive. As a result, most evaluation of off targets is through assessment of binding/activation to accessible normal tissue, cell lines that do not express the target gene and a small number of synthesized peptides designed to elucidate where the therapeutic agent binds to the peptide. These relatively low-throughput and information-poor approaches cannot approach the breadth of possible off targets encoded in the human proteome.

We hypothesized that we could identify large numbers of off-target MHC-I ligands by screening reactivity of TCR like therapeutic molecules to cells expressing DNA barcoded, peptide-minigenes. Building off of the work of others who have used antigen minigenes to express MHC-I ligands in mammalian cells[45-47], we developed a minigene vector called “PresentER” that encodes MHC-I ligands downstream of an endoplasmic reticulum signal sequence. We chose a long signal sequence in order to avoid nonsense-mediated decay of the short messenger RNA. By cloning the precise MHC-I ligand downstream of an endoplasmic signal sequence, the minigene encoded peptides do not need to be cleaved by the proteasome, transported by TAP or trimmed by aminopeptidases in order to be loaded onto MHC-I and presented on the surface of the cells. MHC-I peptides are short (8-11 amino acids), thus the DNA encoding the peptide serves as the barcode which can be re-identified by sequencing the minigene itself.

We have previously isolated and characterized two TCR mimic (TCRm) antibodies (ESK1[30] and Pr20[26]), which bind to the HLA-A*02:01 ligands RMFPNAPYL (WT1:126-134), and ALYVDSLFFL (PRAME:300-309), respectively. WT1 and PRAME are over-expressed across many cancer types but not expressed widely on normal tissue. WT1 is expressed in some developing tissues during embryogenesis, as well as in the adult urogenital system central nervous system and the hematopoietic system[52]. PRAME is also expressed in some normal tissues such as placenta, adrenal glands, ovaries and endometrial lining[53]. Despite this, WT1 remains one of the top National Cancer Institute targets for cancer therapy[48] and PRAME remains a major target for investigational cancer therapeutics[55]. TCR mimic antibodies against WT1 and PRAME were selected by panning a pool of phage displaying the variable region of antibodies called Fragment antigen binding (Fab) against soluble MHC-I complexes containing RMF or ALY peptides. The two TCR mimics that were selected were cloned into monoclonal antibody expression vectors and the purified monoclonal antibody underwent extensive pre-clinical testing to test for specificity to the target epitopes. Cross-reactive ligands of both ESK1 and Pr20 were discovered through alanine scanning and evaluation of a small number of candidate off-target peptides[26,49]. In order to identify if any cross-reactive ligands of ESK1 or Pr20 would compromise the ability to use these drugs in humans, we screened thousands of peptides found in the human proteome to identify those that might bind to ESK1 or Pr20.

Materials and Methods

Cloning PresentER Cassette and PresentER constructs

The Mouse Stem Cell Virus vector that is the basis of the PresentER cassette was a generous gift from Scott Lowe of the Lowe lab. The 98-amino acid ENV_MMTVC signal peptide from Mouse Mammary Tumor Virus envelope protein (accession #Q85646) was found in the Signal Peptide Database (http://www.signalpeptide.de/index.php). IDT gBlocks encoding a modified, human codon optimized signal sequence followed by an antigen and a stop codon were ordered and cloned into the MLP vector with XhoI and EcoRI. The modifications changed two amino acids to enable directional cloning with the SfiI restriction enzyme: …PQTSLTLFLALL[**S>A**]VL[**G>A**]PPPVSG. A cassette with an SfiI site at the 3' end was also included in the gBlock and this construct is collectively termed the PresentER Cassette. In order to clone antigens into PresentER, DNA sequences encoding individual antigens were ordered from IDT, amplified with PresentER-F and PresentER-R primers and digested with SfiI. Digested inserts were purified with the Qiagen MinElute kit to remove primer dimers. The PresentER Cassette was digested with SfiI, treated with Calf Intestinal Phosphatase and gel purified. The inserts were ligated into the digested PresentER backbone with T4 ligase (NEB Catalog #M0202) and transformed into NEB Stable cells (NEB Catalog #C3040). All sequences can be found in the table of DNA and protein sequences. The PresentER cassette and several example PresentER minigenes are available on Addgene (e.g. Plasmid #102944, #102945, #102943).

Production of retrovirus and transduction of cells

HEK293T Phoenix amphoteric cells were transfected with polyethylenimine (PEI) and PresentER plasmid (15µg DNA : 45µg PEI) in 10cm TC plates. Virus was harvested every 12 hours, pooled, concentrated with Clontech Retro-X (Cat. #631455) and frozen in aliquots. T2 cells were spinoculated with virus in non-TC treated 6-well plates at 32ºC x 2,000xg for 2 hours with 4µg/ml of polybrene. RMA-S cells were spinoculated with virus in non-TC treated 6-well plates at 32ºC x 1,000xg for 2 hours with 4µg/ml of polybrene. Library retrovirus was produced in the same way, except that virus production was scaled up to four 15cm plates per library. The volume of viral supernatant that led to 1/3 maximal transduction efficiency was established for each batch of virus produced. Transduced cells were selected with 1µg/ml (T2) or 4µg/ml (RMA-S) of puromycin for 2-3 days until cells were >95% GFP positive.

Antibodies and commercial reagents

ESK1 and Pr20 antibodies were purified by Eureka Therapeutics and fluorescently labeled using Innova Biosciences lightning link kits (705-0010). Each labeled aliquot was titered on T2s or RMA-S cell bearing the PresentER minigenes. The TCR multimer specific to NLVPMVATV (CMV aa495-503)/HLA-A2.1 was purchased from Altor BioScience (Cat #TCR-CR1-0020). APC labeled antibodies specific to SIINFEKL/H2-Kb (Clone 25-D1.16) were purchased from Ebioscience (Cat #141606).

Generation of AviTagged A6 TCR

The A6 TCR was a generous gift from Brian Baker's lab. The beta chain plasmid was modified to encode a C-terminal AviTag biotinylation site (GLNDIFEAQKIEWHE). The site was inserted using Gibson cloning and following two primers: F: 5'-gcagaaaattgaatggcatgaaTAAGCTTGAATTCCGATCCGG-3' R: 5'-gcttcaaaaatatcgttcaggccGTCTGCTCTACCCCAGGC-3'. Both the alpha and beta

chain were expressed in BL21 (DE3) bacteria and induced with 1mM IPTG. The beta chain was co-expressed with a plasmid encoding the BirA enzyme (Addgene #26624) and supplemental biotin (0.5mM D-Biotin). Inclusion bodies were harvested and the individual chains were purified and re-folded together according to previously described protocols[50,51].

Cloning the PresentER Library

A pool of 12,472 oligonucleotides was synthesized by CustomArray, Inc in the following format: 5'- GGCCGTATTGGCCCCGCCACCTGTGAGCGGG…[27-30nt insert]…TAAGGCCAAACAGGCC-3'. The oligonucleotides were cloned into the PresentER vector in exactly the same way as the individual minigene. After ligation, the ligation products were electroporated into competent cells and plated onto four 15cm ampicillin-LB plates. After overnight growth, the number of colonies was estimated to be ~$46x10^6$. The colonies were scraped off the plate and grown for 3.5h in TB + ampicillin at 37ºC at 225rpm. The bacteria were maxiprepped and library representation was checked by Illumina sequencing.

Library screening by FACS

T2 cells transduced with library virus or single-minigene controls were stained with ESK1-APC or Pr20-APC. Sorting gates were set-up based on the fluorescence of the stained single-minigene control cells (PresentER-RMF and PresentER-ALY). Each sort was performed two times for each antibody. Sorted cells were frozen for genomic DNA extraction and Illumina sequencing. Pre-sort ("unsorted") cells were frozen as well.

Genomic DNA Extraction and Library Sequencing

Genomic DNA was extracted from unsorted or sorted cells with the Qiagen Gentra Puregene kit (Catalog #158667). Genomic DNA was amplified with barcoded primers and sequenced by the Integrated Genomics Operation core at MSKCC.

Bioinformatics

The peptides included in the PresentER library were found in Uniprot TrEMBL database of reviewed and unreviewed human protein sequences. Substrings of unique 9 and 10 amino acid sequences were derived from this databaset and affinity to HLA-A*02:01 was calculated using NetMHCPan. Peptides with predicted ic50<500nM were compared to the ESK1 and Pr20 cross-reactivity motifs to determine which should be included in the library (**Figure 14**). All potentially cross-reactive ESK1 ligands and half of the potential Pr20 ligands were included in the library.

After Illumina sequencing, reads were mapped to the PresentER minigene library with Bowtie2. Reads that did not map to the minigenes in the library were discarded. Data analysis was performed in R. In order to identify minigenes encoding antibody ligands, we normalized the abundances in the sorted samples by the abundances in the unsorted samples and then divided the abundance of each minigene in the "antibody high" sample by the abundance of each minigene in the "antibody low" sample.

Screen validation by peptide pulsing

Spot synthesized peptides (PepTrack libraries) were ordered from JPT Peptide Technologies. Peptides were resuspended to 20mg/ml in DMSO, followed by dilution

to 1mg/ml in PBS and finally pulsed onto T2 cells at 50µg/ml. Pulsed cells were stained with ESK1 or Pr20 to evaluate binding of antibody to pMHC.

Evaluation of ESK1 binding to JY and TPC1

JY and TPC1 cells were stained with unlabeled ESK1 or IgG1 isotype control (BioLegend Cat #ET901) followed by anti-human IgG1 APC antibody (BioLegend Cat #HP6017).

Generation of T cells with transgenic TCR

MSGV-1 plasmids encoding the DMF5 and 1G4 TCRs were provided as a kind gift from Dr. Steven A. Rosenberg. Plasmids encoding A6 and B7 TCRS were generated as described above. TCR retroviral transduction was performed as described previously[52]. Briefly, retroviral particles were generated by transient transfection of the retroviral packaging cell line 293GP cells with the pMSGV1-TCR plasmids and pRD114 plasmid using Lipofectamine 2000 (Life Technologies). Retroviral supernatant was harvested 2 days later, and used to transduce PBMC that were stimulated with soluble 50 ng/ml anti-CD3 (OKT3, Miltenyi Biotec) and 300 IU/ml rhIL-2 (Chiron) for 2 days prior to retroviral transduction. Retroviral transductions were performed on Retronectin (Takara) coated non-tissue culture treated 24-wells plates by spinoculation of the retrovirus at 2,000× g, 32°C for 2 hours, followed by addition of activated T cells to the retrovirus containing plates. Following overnight incubation at 37°C, 5% CO2, T cells were transferred to a tissue-culture treated 24-wells plate and expanded in T cell media supplemented with 300 IU/ml rhIL-2 (Chiron). Transduced T cells were used at 10-15 days post-transduction or cryopreserved until used in assays.

Generation of A6 and B7 TCR mammalian expression plasmids

The bacterial expression vectors for production of non-membrane bound (soluble) A6 and B7 TCRs were cloned with Gibson Assembly into the MSGV-1 mammalian expression (retroviral) vector. MSGV-1 vector encoding 1G4 was used as the vector backbone and the intergenic sequence between the 1G4 alpha and beta chains were cloned in between the alpha and beta chains of A6 and B7. A gBlock encoding the common N-terminus of A6, B7 and 1G4 was ordered (MKSLRVLLVILWLQLSWVWSQ): 5'-CGCAGCTTGGATACACGCCGCCCACGTGAAGGCTGCCGACCCCGGGGGTGGACCATCCTCTAGACCGCCATGAAGTCTTTGCGCGTACTCTTGGTGATATTGTGGCTCCAATTGAGTTGGGTGTGGTCCCAG-3'. An additional 2 gBlocks were ordered for the A6 beta chain (5'-CGGCCAGGTCTTGCCGGGGGACGACCAGAGCAGTATTTCGGGCCAGGGACGCGCCTTACGGTAACAGAAGACTTGAAGAATGTCTTTCCACCTGAGGTCGCCGTTTTTGAACCCTCCGAGGCCGAAATAAGTCATACTCAAAAGGCGACTCTGGTGTGCCTCGCCACCGGGTTTTACCCGGACCACGTAGAACTTAGCTGGTGGGTGAATGGTAAAGAGGTCCATAGCGGGGTGTGCACGGACCCACAGCCTCTCAAGGAACAACCCGCTCTGAATGATTCCAGGTATTGTCTTAGCTCACGGCTTCGAGTGTCAGCTACTTTTTGGCAAGAT-3' and B7 beta chain (5'-AGTTACCCTGGAGGAGGCTTTTATGAGCAGTATTTCGGTCCTGGAACAAGGCTGACCGTGACGGAAGATTTGAAAAATGTCTTTCCCCCAGAGGTAGCAGTCTTCGAGCCGTCCGAGGCCGAGATATCCCATACCCAGAAGGCAACCCTTGTTTGCTTGGCAACGGGATTTTATCCAGATCATGTGGAATTGTCCTGGTGGGTCAACGGCAAAGAGGTTCACAGCGGCGTCTGCACAGATCCGCAACCACTCAAGGAACAGCCCGCTCTTAATGATTCTCGCTACTGTCTGAGTTCCAGGTTGCGGGTCAGCGCTACTTTCTGGCAGGAT-3'): The following fragments were PCR amplified and then assembled:

#	Fragment name	Template DNA	F Primer Name	F primer sequence	R primer Name	R primer sequence
A1	N terminus of A6 Alpha chain	A6B71G4 Alpha chain N terminal gblock	A6B7Oligo2	ccctcaaagtagacggcatcCGCAGCTTGGATACACGC	A6Oligo3	cttccttctgCTGGGACCACACCCAACTC
B1	N terminus of B7 Alpha chain	A6B71G4 Alpha chain N terminal gblock	A6B7Oligo2	ccctcaaagtagacggcatcCGCAGCTTGGATACACGC	B7Oligo3	cattcttctgttgCTGGGACCACACCCAACTC
A2	A6 alpha chain	A6 alpha	A6Oligo4	gtggtcccagCAGAAGGAAGTGGAGCAG	A6Oligo5	tgacatcacaGGAACTTTCTGGGCTGGG
B2	B7 alpha chain	B7 alpha	B7Oligo4	gtggtcccagCAACAGAAGAATGATGACCAGCAAG	B7Oligo5	tgacatcacaAGAGCTTTCCGGGCTCGG
A3	pGMT7 region in between chains for A6	1G4	A6Oligo6	agaaagttccTGTGATGTCAAGCTGGTC	A6Oligo7	gacctggccgGCTGGCACAGAAGTACAC
B3	pGMT7 region in between chains for B7	1G4	B7Oligo6	ggaaagctctTGTGATGTCAAGCTGGTC	B7Oligo7	cagggtaactGCTGGCACAGAAGTACAC
A4	A6 beta chain fragment	gBlock A6betachainfrag	A6Oligo8	ctgtgccagcCGGCCAGGTCTTGCCGGG	A6Oligo9	cagcggaagtggttgcggggATCTTGCCAAAAAGTAGCTGACACTCGAAGC
B4	B7 beta chain fragment	gBlock B7betachainfrag	B7Oligo9	ctgtgccagcAGTTACCCTGGAGGAGGC	B7Oligo10	cagcggaagtggttgcggggATCCTGCCAGAAAGTAGC
5	Backbone	1G4	A6B7Oligo1	GATGCCGTCTACTTTGAG	A6B7Oligo10	CCCCGCAACCACTTCCGC

ELISPOT and co-culture assay

IFN-gamma release ELISPOTs were performed in 200µl of RPMI supplemented with 5% FBS. 50,000 transduced T cells were incubated at the indicated effector:target (E:T) ratios (typically 1:1) with T2 cells expressing PresentER peptides as targets. Co-culture assays were performed in a similar manner: 50,000 T cells were incubated with 50,000 target cells per well of a U-bottom plate and the remaining cells were analyzed by flow cytometry to evaluate specific lysis of target cells.

Data Availability

Sequencing data for each of the experiments has been deposited:

Minigene sequencing of sorted T2 cells expressing a library of HLA-A*02:01 exomic peptides. Cells are sorted for high and low binding of the TCR mimic antibody ESK1. DOI:10.5281/zenodo.1313110
Minigene sequencing of sorted T2 cells expressing a library of HLA-A*02:01 exomic peptides. Cells are sorted for high and low binding of the TCR mimic antibody Pr20 DOI:10.5281/zenodo.1326544
Minigene sequencing of T2 cells expressing a library of off targets (derived from A6 and B7 binding motifs in Hausmann 1999) after co-culture with A6, DMF5 or 1G4 expressing T cells (from a non-A2 donor) DOI:10.5281/zenodo.1341943
Minigene sequencing of T2 cells expressing a library of off targets (derived from A6 and B7 binding motifs in Hausmann 1999) are co-cultured with A6, B7, DMF5 or 1G4 expressing T cells (from a non-A2 donor) DOI:10.5281/zenodo.1342624

Results

PresentER minigene design and validation

The PresentER minigene was designed in two stages. First, we cloned into the MLP vector an endoplasmic reticulum (ER) signal sequence from the Mouse mammary tumor virus (MMTV). This sequence was modified at the 3' end to include a removable cassette surrounded by SfiI restriction sites where DNA sequences encoding MHC-I ligands can be cloned (**Figure 6**). This vector is called the "PresentER Cassette" and includes an eGFP fluorescent marker and a Puromycin resistance gene. Subsequently, oligonucleotides encoding the HLA-A*02:01 ligands RMFPNAPYL (Human WT1:126-134), and ALYVDSLFFL (Human PRAME:300-309) were cloned into the SfiI-bracketed region of the PresentER vector. Tap deficient HLA-A*02:01^{+} T2 cells transduced with the PresentER-ALY and PresentER-RMF minigenes were stained with fluorescently labeled ESK1 and Pr20 to evaluate binding to MHC-I ligands expressed by the PresentER vector. Both ESK1 and Pr20 weakly bound T2 cells expressing their respective PresentER epitopes, but not irrelevant epitopes **(Figure 7A-B)**. In order to demonstrate that PresentER encoded ligands could be presented on non-HLA-A*02:01 alleles, we cloned SIINFEKL (Chicken Ovalbumin:257-264) and MSIIFFLPL (mouse PEDF:271-279) into PresentER and showed that a TCR mimic antibody specific to H2-Kb/SIINFEKL (clone #25-D1.16) bound only to PresentER-SIINFEKL expressing Tap deficient RMA/S cells (**Figure 7C)**.

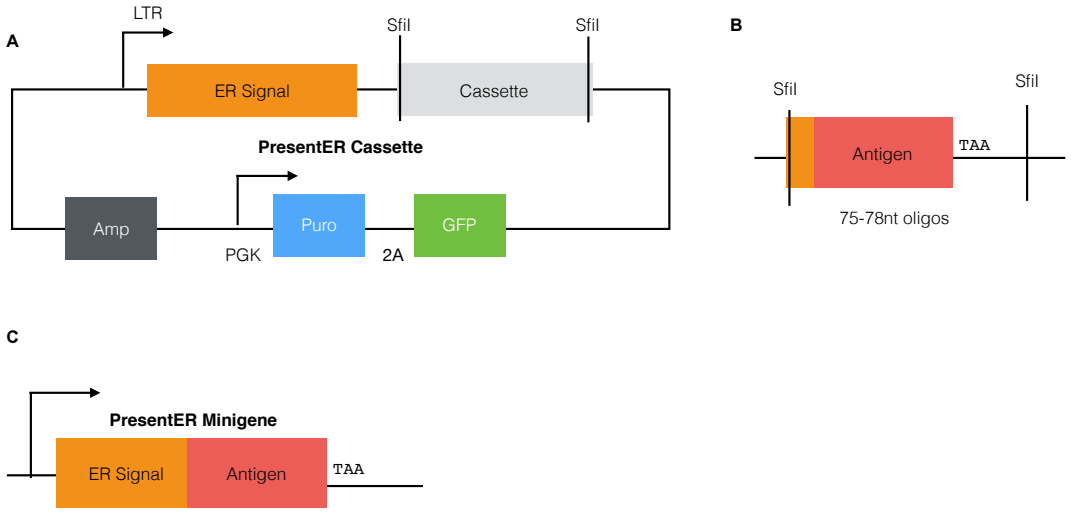


Figure 6 Schematic of PresentER cassette and minigene. (A) The PresentER vector is based on an MSCV retroviral vector. The peptide antigen minigene is driven by the MSCV LTR and encodes an endoplasmic reticulum (ER) targeting sequencing followed by the precise peptide to be expressed and a stop codon. The vector contains a puromycin resistance gene and eGFP driven by the PGK promoter. The ER sequence is the leader sequence from MMTV gp70 protein. A removable cassette (for easy cloning) is bounded by SfiI restriction sites. **(B)** The oligonucleotide sequence encoding the antigen is an inexpensive 75-78nt sequence and can be amplified, digested with SfiI and ligated into the backbone. **(C)** A schematic of the final PresentER minigene construct is shown.

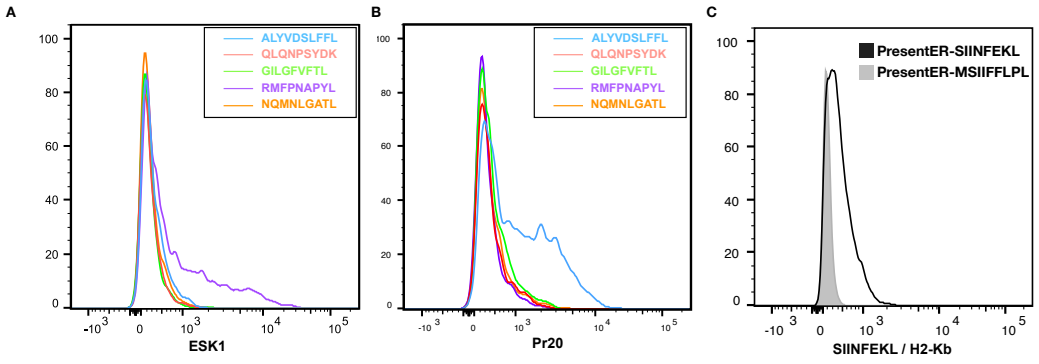


Figure 7: Binding of TCR mimic antibodies to human and mouse cell lines expressing PresentER antigens. T2 cells expressing PresentER-RMFPNAPYL (purple) and PresentER-ALYVDSLFFL (blue) are bound by fluorescently labeled **(A)** ESK1 and **(B)** Pr20, respectively. Irrelevant HLA-A*02:01 ligands serve as negative binding controls: Influenza M peptide 58-66 (green), Human WT1 239-247 (orange) and Human EW QLQNPSYDK (red). **(C)** A fluorescently labeled antibody to SIINFEKL/H-2Kb (clone 25-D1.16) binds to RMA-S cells expressing PresentER-SIINFEKL, but not to the H-2Kb ligand PresentER-MSIIFFLPL.

In order to verify that TCRs could bind to the presented peptides, we first obtained or purified two fluorescently labeled TCR multimers and verified that they could bind to their cognate targets. T2 cells expressing PresentER-NLVPMVATV (Cytomegalovirus pp65:495-503; HLA-A*02:01 ligand) were specifically, but weakly, bound by a high affinity TCR multimer directed to this epitope (**Figure 8A**). A tetramer made of A6 TCR molecules linked to R-phycoerythrin (R-PE) showed specific binding to T2 cells expressing its target MHC-I ligand PresentER LLFGYPVYV (HTLV-1 Tax:11-19; HLA-A*02:01 ligand) **(Figure 8B).**

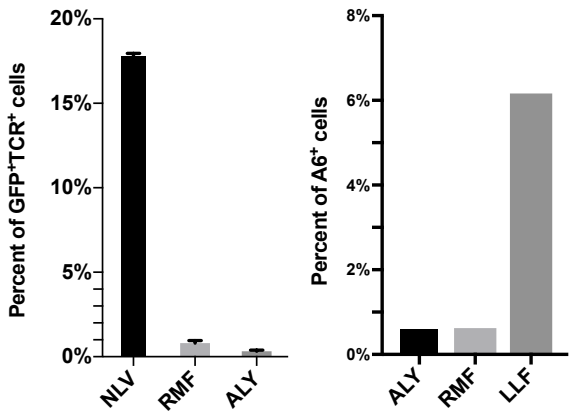


Figure 8 Binding of TCR multimers to T2 cells expressing PresentER antigens. **(A)** T2 cells expressing PresentER-NLVPMVATV 495-503 were bound by a fluorescently labeled STAR CMV pp65 directed TCR multimer. **(B)** A fluorescently labeled A6 TCR tetramer bound to its target ligand, HTLV-1 Tax 11-19 LLFGYPVYV. Irrelevant HLA-A*02:01 ligands serve as negative binding controls.

In order to verify that presented peptides could be properly seen by T cells and thus that the system could be used for functional immunologic assays, we obtained two vectors encoding mammalian TCR that recognize two different cancer-associated antigens. The DMF5 TCR recognizes MART-1 27-35 (AAGIGILTV) while the 1G4

TCR recognizes NY-ESO-1 157-165 (SLLMWITQC). In addition, we cloned the A6 and B7 TCRs—which both recognize HTLV-1 Tax 11-19 (LLFGYPVYV)—out of the bacterial expression constructs (which lack intracellular domains) and into retroviral mammalian expression constructs. Human donor lymphocytes were transduced with the four TCRs and the ability of the T cells to activate and kill T2 PresentER minigene expressing target cells in co-culture was assayed by IFNg release ELISpot (**Figure 9**) and by killing of target cells (**Figure 10A-B**). T cells expressing the A6, 1G4 and DMF5 TCRs consistently activated at high levels when challenged with T2 cells expressing their cognate epitope. However, T cells expressing the B7 TCR consistently yielded low levels of activation. Activated splenocytes from OT-1 mice, which express a transgenic TCR specific to the SIINFEKL peptide (Ova 257-264) in the context of H-2Kb, were also able to kill co-cultured RMA/S cells expressing PresentER-SIINFEKL minigenes, but not irrelevant control minigenes (**Figure 10C**).

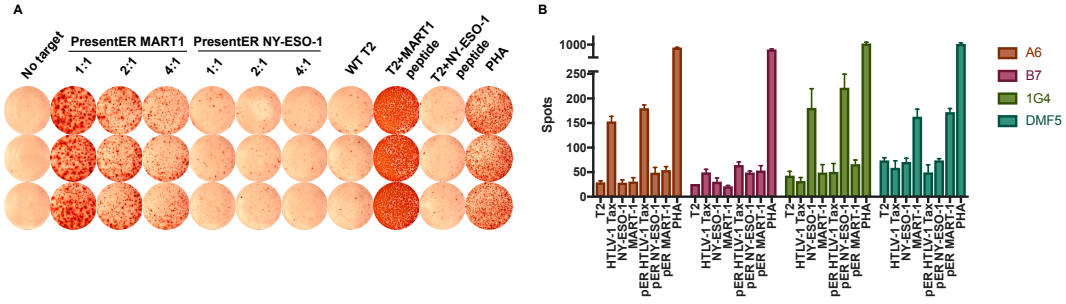


Figure 9 T cell activation by PresentER expressing target cells. (A) ELISpot of genetically engineered T cells expressing the DMF5 TCR directed against the HLA-A*02:01 ligand MART-1 27-35 (AAGIGILTV) challenged with T2 cells pulsed with peptide, expressing the PresentER MART-1 minigene or irrelevant NY-ESO-1 157-165 (SLLMWITQC) peptide. **(B)** Genetically engineered T cells expressing the DMF5, 1G4, A6 or B7 TCRs were challenged with peptide pulsed T2s and T2s expressing PresentER antigen minigenes. Quantification of spots is presented.

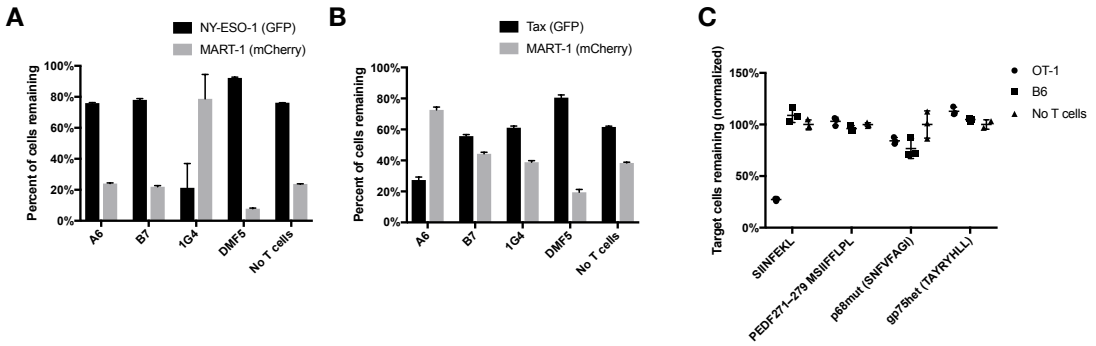


Figure 10 T cells kill co-cultured PresentER expressing target cells. 96-hour co-culture killing assays *in vitro* where **(A-B)** A6, B7, 1G4 and DMF5 expressing T cells were incubated with T2 PresentER MART-1 (mCherry) cells mixed with **(A)** T2 PresentER NY-ESO-1 minigene (GFP) cells or **(B)** Tax (GFP) cells and surviving target cells were quantified by flow cytometry. The fraction of target cells is normalized to the percentage in the wells without T cells. (**C)** RMA/S cells expressing SIINFEKL, MSIIFFLPL, SNFVFAGI or TAYRYHLL minigenes were mixed with activated splenocytes from OT-1 mice, C57BL6/N mice or no T cells. The fraction of target cells is normalized to the percentage in the wells without T cells.

Based on the fact that PresentER drives antigen presentation in the absence of the TAP peptide transport, we hypothesized that the PresentER system would bypass proteasomal cleavage and processing in the generation of peptide-MHC. To test whether any peptide processing was occurring, we immunoprecipitated peptide-MHC complexes from T2 cells expressing PresentER-RMF or PresentER-ALY and identified bound peptides by mass spectrometry. RMFPNAPYL and ALYVDSLFFL were identified only in cells encoding those PresentER constructs (**Figure 11A-B**). Truncated versions of the RMF and ALY peptides were not found, nor were peptides derived from the ER signal sequence. In order to demonstrate that peptide presentation was dependent on the endoplasmic reticulum targeting sequence and not via another transport mechanism, we scrambled the ER signal sequence of the PresentER-RMF and PresentER-ALY constructs and stained them with ESK1 and Pr20. We found no binding to cells expressing scrambled constructs, indicating that the signal sequence is required for presentation of peptides in T2 cells **(Figure 11C)**.

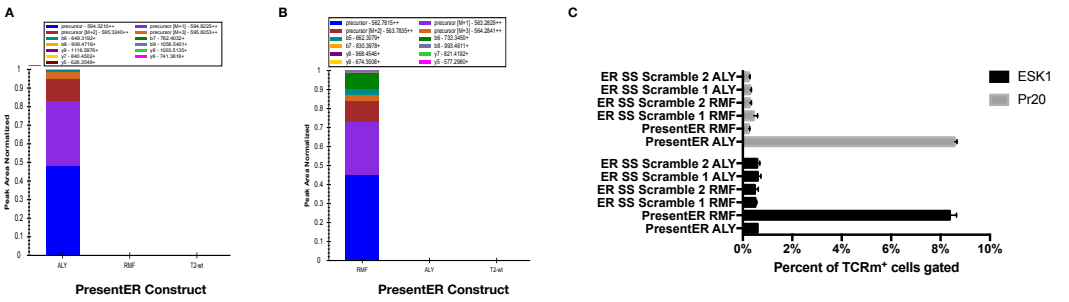


Figure 11 MHC-I ligands produced by PresentER minigenes are specific to the encoded peptide and depend on the ER signal sequence. MHC-I ligands were isolated from wild-type T2 cells or T2 cells expressing **(A)** PresentER-ALY or **(B)** PresentER-RMF. Peptides were identified by mass spectrometry and the intensity of ions corresponding to the ALY (**A**) or RMF (**B**) peptides in each sample are plotted. (**C**) ESK1 and Pr20 binding to T2 cells transduced with PresentER-RMF and PresentER-ALY is higher compared with binding to T2 cells expressing constructs with scrambled signal sequences.

These experiments demonstrate that PresentER minigenes yield functional, mature MHC-I antigens which can be bound by TCR like molecules and recognized by T cells. Binding of fluorescently labeled TCR like molecules to cells expressing their target antigens using PresentER is significantly weaker than compared to binding to pulsed cells, which is expected given the significantly lower amount of antigen yielded by endogenous antigen presentation as compared to pulsing with micromolar quantities of soluble peptide. However, the levels of T cell activation between PresentER encoded antigens and pulsed peptides are similar. This is consistent with the fact that T cells require very few target pMHC in order to activate a productive T cell response[53,54]. Nevertheless, we proceeded to see if libraries of PresentER antigens could be used to screen for off targets of TCR like monoclonal antibodies.

If the PresentER system is to dramatically increase throughput in the identification of cross-reactive targets of TCRs, it must be usable in a pooled screen, which is a form of high throughput genetic screening. In classical high throughput screening (HTS),

large numbers of chemical treatments or genetic variants are tested, one-by-one, and treatments/genes which yield a desired phenotype are identified. These screens are typically performed in 96 or 384 well plate format so that at each well in the plate it is known which treatment was used. Pooled screens enable less costly HTS of genetic variants because, instead of synthesizing individual genetic constructs and testing each gene in a separate well, all the genetic variants are cloned simultaneously in one reaction (typically into a retroviral vector) and a population of cells are transduced such that each cell expresses only one genetic variant on average. Attaining single-variant-per-cell transduction is typically achieved by low multiplicities of infection (MOI) such that, on average, no cell receives more than one copy of a genetic construct and most cells receive no constructs. Cells which have not received a genetic construct are negatively selected. However, a major caveat of pooled screens is that the genetic constructs must be potent enough that a single genetic copy is capable of driving the phenotype[55]. In contrast to traditional HTS where cells may receive many copies of a gene, and thus single-copy competence is not as important, pooled screens fail unless a single copy of the genetic variant is sufficient for screening. To demonstrate that the PresentER system is single-copy competent, we transduced T2 cells with PresentER-RMF and PresentER-ALY with a 5-fold range of viral supernatant volume and achieved a 1-log difference in percentage of infected cells between the highest and lowest MOI transductions. Despite a 10-fold difference in the number of infected cells, the GFP positive cells in the population drive the same level of antigen presentation, thereby demonstrating that the system is single-copy competent and can be used for pooled screens **(Figure 12)**.

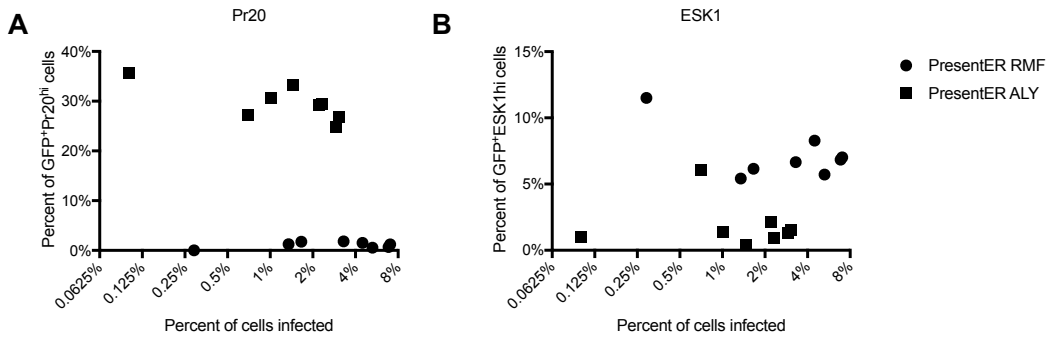


Figure 12 PresentER is a single-copy competent MHC-I antigen expression system. T2 cells were transduced with 5-fold serially diluted PresentER-RMF or PresentER-ALY viral supernatant in triplicate for each volume of virus. The percent of GFP positive, ESK1 positive cells after transduction is plotted as a function of percentage of cells infected (**B and D**). Across a 1-log range of low MOI infections, the same percentage of ESK1^{hi} and Pr20^{hi} cells are noted, indicating that a single copy of the minigene must be capable of driving antigen presentation.

A PresentER screen of ESK1 reveals hundreds of cross-reactive pMHC

During the initial characterization of both Pr20 and ESK1, low-throughput alanine/residue scanning identified several off-target HLA-A*02:01 ligand peptides. In combination with a crystal structure of ESK1 bound to RMFPNAPYL[49] (**Figure 13**), we determined that ESK1 binding to RMFPNAPYL depended primarily on the R1 and P4 residues. In contrast, Pr20 bound mostly to the C-terminus of the ALYVDSLFFL peptide[26]. Therefore, we constructed a biased library ("PresentER Library #1") of possible ESK1 and Pr20 cross-reactive targets by searching the human proteome *in silico* for 9-mer and 10-mer peptide sequences that might be off targets. For ESK1, we searched for peptides containing arginine in position 1 and proline in position 4. We performed a similar search for Pr20 cross-reactive targets using a 10-mer peptide motif based on prior biochemical data **(Figure 14a)**. We located 1,157 and 24,539 potential cross-reactive peptides of ESK1 and Pr20, respectively, with NetMHCPan[56] predicted HLA-A*02:01 affinity of less than 500nM. We synthesized a pool of 12,472

oligonucleotides that together encoded all of the ESK1 cross reactive peptides and half of the Pr20 cross-reactive targets plus the one amino acid mutants of RMF and ALY (termed “CR-ESK1” and “CR-Pr20”, respectively), as well as positive/negative controls **(Figure 14b)**. The oligonucleotide library was cloned into the PresentER vector following a scaled-up protocol of the cloning of individual minigenes. To ensure that exponential amplification during PCR had no compromised the distribution of minigenes in the population, the cloned library was sequenced by Illumina HiSeq. All synthesized minigenes were found in the plasmid pool and no extreme outliers were noted **(Figure 15)**.

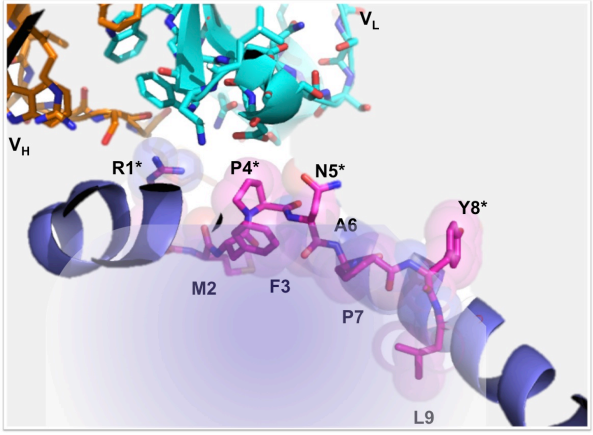


Figure 13 Crystal structure of ESK1 in complex with HLA-A*02:01/RMFPNAPYL. RMFPNAPYL binds HLA-A*02:01 with Met2, Phe3, Pro7, and Leu9 facing the HLA receptor and Arg1, Pro4, Asn5, and Tyr8 facing ESK1. The heavy and light chains of ESK1 are in orange and cyan. Arg1, Pro4 and Asn5 are within van der Waals distance to ESK1. Figure and caption adapted from Ataie et al 2018[49],

A

	1	2	3	4	5	6	7	8	9	10
ESK1 Target	R	M	F	P	N	A	P	Y	L	–
ESK1 Library	R	*	*	P	*	*	*	*	*	–
Pr20 Target	A	L	Y	V	D	S	L	F	F	L
Pr20 Library	*	*	*	DE	KH R	K H R	K H R	FW YV LI	FW YV LI	L V I

B

		# Constructs (%)
	Controls	13 (0.1%)
CR-ESK1	RMFPNAPYL: all 1 AA mismatch	180 (1.45%)
	ESK1 genomic off-targets	1,157 (9.3%)
CR-Pr20	ALYVDSLFFL: all 1 AA mismatch	190 (1.5%)
	Pr20 genomic off-targets	10,893 of 24,539 (87.6%)
	Total	**12,433**

Figure 14 Design of PresentER library #1 and ESK1 screen.
(A) Design of sequence constraints on the target peptide library based on prior biochemical data. The human exome was mined for peptides matching the specified ESK1 and Pr20 consensus motifs. Asterisks indicate any amino acid is allowed. Red characters indicate prohibited amino acids at that position and black characters indicate allowed amino acids at that position. **(B)** The constructed library included 13 control peptides, 1,337 CR-ESK1 peptides and 11,083 CR-Pr20 peptides.

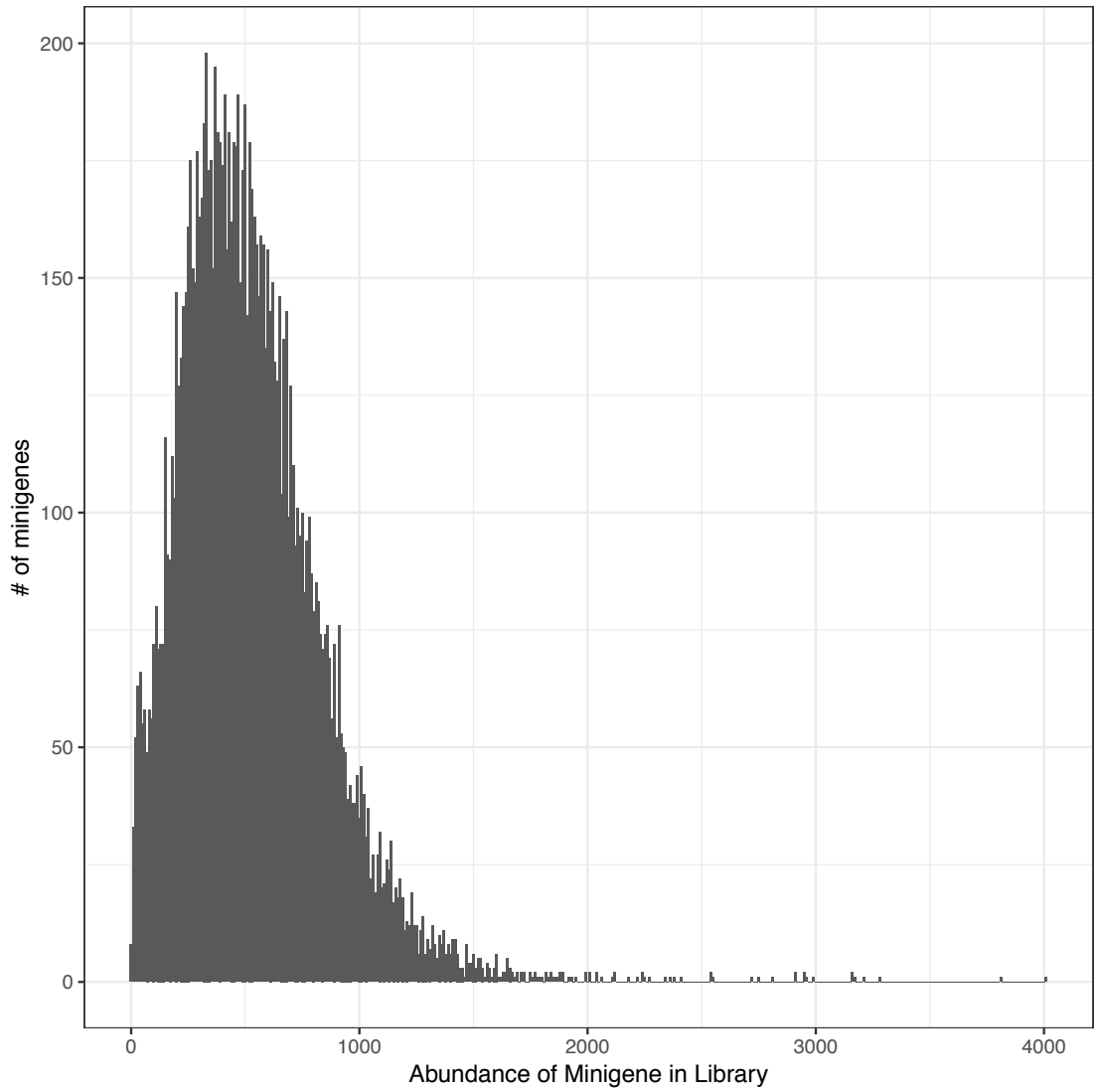


Figure 15 PresentER Library #1 minigene representation analysis.
Minigene sequences from PresentER Library #1 were amplified and sequenced by Illumina HiSeq. All synthesized minigenes were located in the cloned library and, while there were some highly abundant outliers, no outlier was present at more than a log higher than the median abundance in the population.

The flow-based screen assaying for cross-reactivity was performed as follows:

HEK293T Phoenix Amphoteric cells were transfected with library plasmid. After 24 hours, viral supernatant was harvested every 12 hours, concentrated using Retro-X-Concentrator, and frozen in aliquots. The library viral supernatant functional titer was

determined under the same transduction conditions as used for transduction of T2 cells in order to determine the volume that yielded 1/3 maximal MOI. T2 cells were transduced at 1/3 MOI and transductants were selected with puromycin. Library representation was kept at >1,000x at all stages. Finally, transduced T2 cells were stained with ESK1 or Pr20 and sorted by fluorescence-activated cell sorting (FACS) into two populations: high binders and low binders (i.e. ESK1^{hi}, ESK1^{low}, Pr20^{hi}, Pr20^{low}. Genomic DNA from sorted cells was extracted, minigenes were amplified by PCR and sequenced by Illumina HiSeq. Sequences were aligned to the library of minigenes and the number of reads mapping to each minigene was quantified. A schematic of the screen is presented in **Figure 16**.

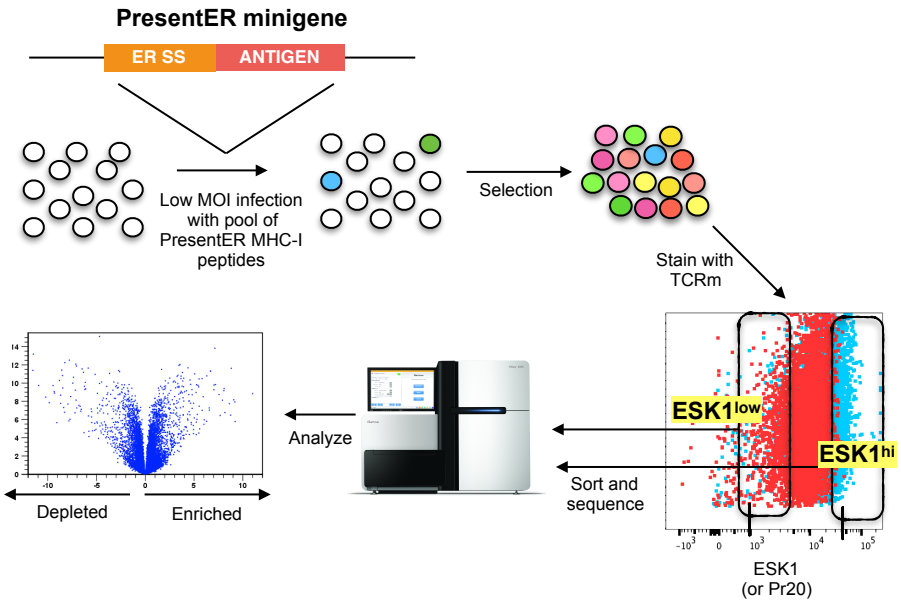


Figure 16 Schematic of the flow-based screen. T2s are transduced at low MOI (<0.3) with retrovirus encoding a pool of PresentER minigenes. Transduced cells are selected by puromycin and then cultured until sufficient cells are obtained. Cells are stained with the TCR mimic antibody and fluorescent activated cell sorting is used to sort binding and non-binding populations of cells. Genomic DNA is extracted from sorted cells and sequenced with Illumina sequencing.

The abundance of each minigene in the unsorted samples was well correlated (r = 0.93), indicating that sample handling and the days between sorts did not affect the library representation **(Figure 17A)**. An ESK1 binding enrichment score was calculated for each minigene, taking into account the abundance of each minigene in the $ESK1^{hi}$, $ESK1^{low}$ and unsorted samples:

(frequency in $ESK1^{hi}$) / (frequency in $ESK1^{low}$) / (frequency in unsorted).

Minigenes encoding previously known ESK1 ligands had higher enrichment scores than those encoding non-ligands (p=0.032), suggesting that the flow-based screen was able to separate ESK1 binders from non-binders (**Figure 18**)

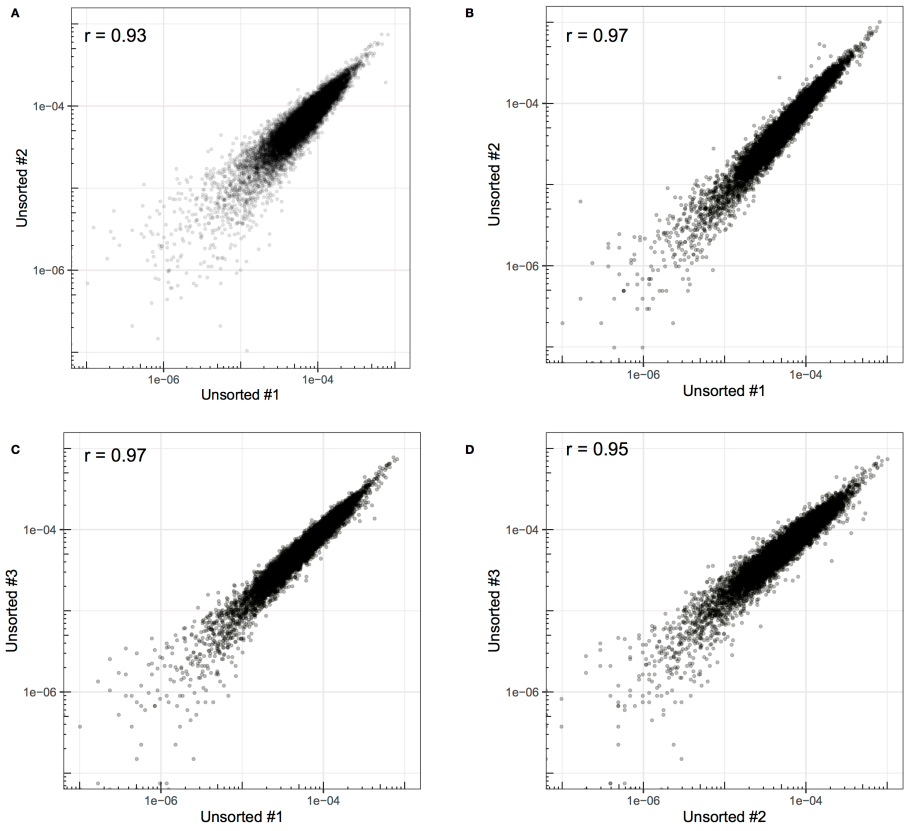


Figure 17 Minigene abundance in unsorted samples is highly correlated. The abundance of each minigene in the two unsorted samples taken before ESK1 sorting **(A)**. The abundance of each minigene in the three unsorted samples taken prior to Pr20 sorting: 1 vs 2 **(B)**, 1 vs 3 **(C)** and 2 vs 3 **(D)**. Correlation coefficients are reported.

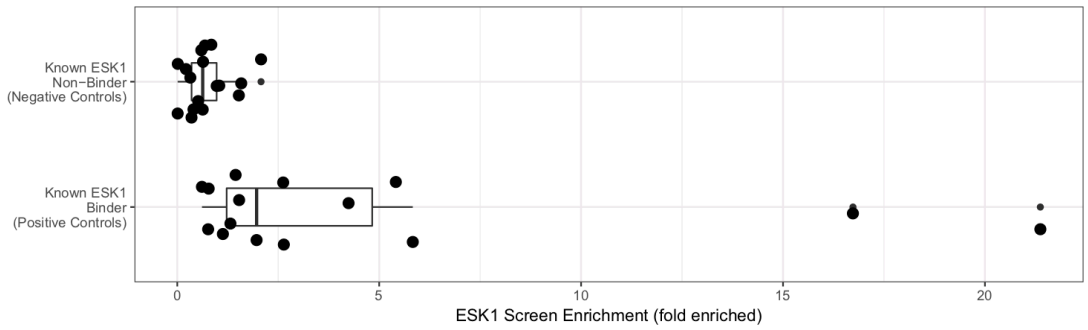


Figure 18 Enrichment of known ESK1 binders in screen. The enrichment scores of control peptides known to be either ESK1 binders or know ESK1 non-binders are compared. Known ESK1 binders were enriched in the pull down (p=0.032).

A scatter plot comparing ESK1 enrichment and NetMHCPan predicted HLA A*02:01 binding affinity of all 12,434 minigenes in the library is presented in **Figure 19**. Surprisingly, several of the most enriched peptides that emerged in the ESK1 screen were CR-Pr20 peptides (marked as squares). Although the target of ESK1 is a 9-mer, several of the CR-Pr20 10-mers that were enriched for ESK1 binding bore sequence similarity to RMFPNAPYL, the target of ESK1, such as RVIMPCNWWV and RMFSGVGVYL. To validate the ESK1 hits, we synthesized soluble peptide for a subset of screen hits and assayed them for binding to ESK1 by pulsing T2 cells. Of the 27 peptides tested, 22 (81%) showed increased binding to ESK1, including several CR-Pr20 peptides which did not even contain a proline in position 4 **(Figure 20)**. Of great importance is that these unusual ESK1 ligands could not have been predicted from either the crystal structure of ESK1 or the alanine scanning data.

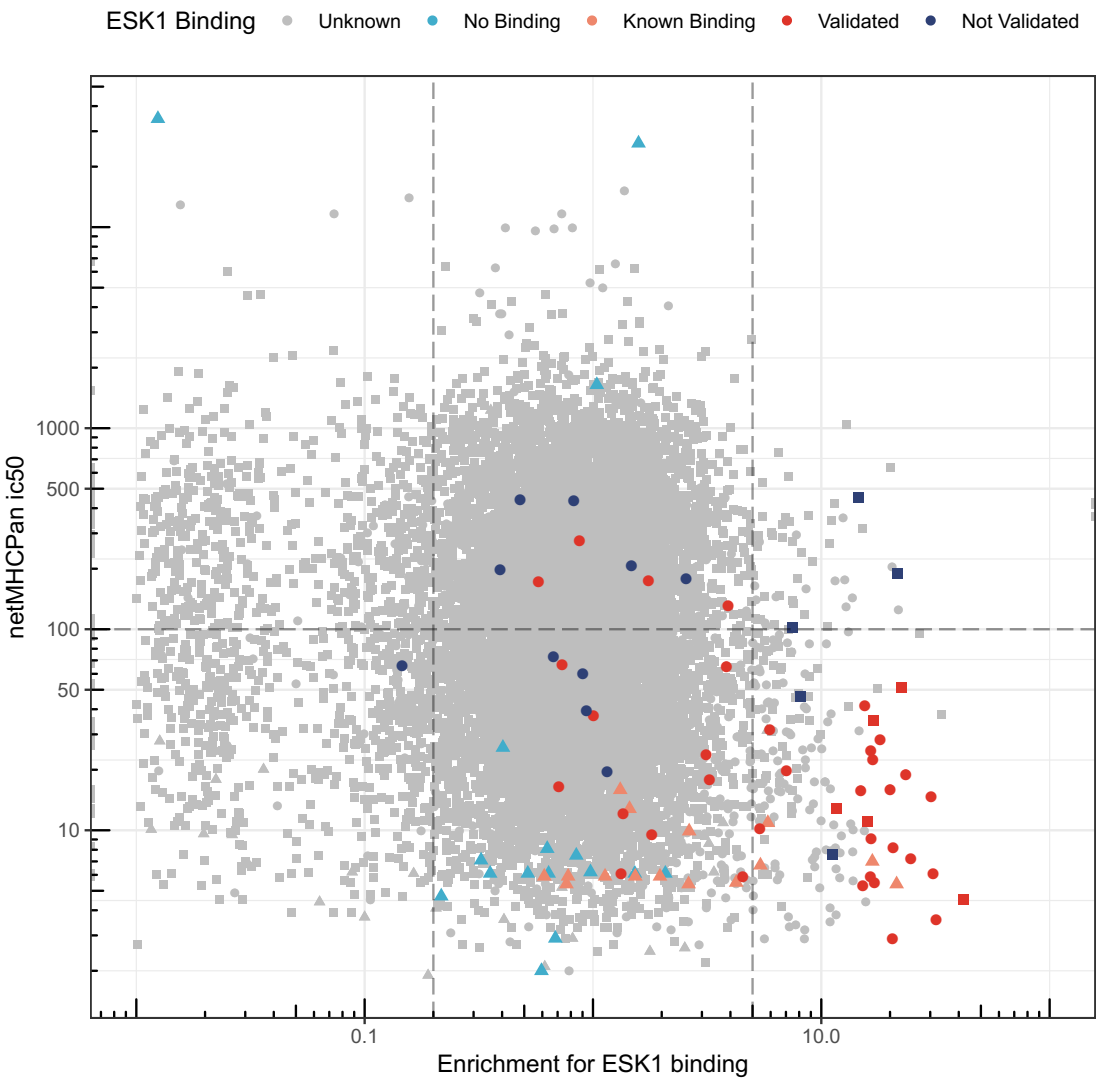


Figure 19 Scatterplot of the ESK1 library screen. Each point is a unique peptide minigene with the x-axis indicating enrichment for ESK1 binding (with 1 set as no enrichment) and y-axis indicating the peptide's predicted affinity (ic50 nM) to HLA-A*02:01. Lower ic50 indicates higher affinity. Marked negative control ESK1 non-binders are marked as blue triangles. Previously known ESK1 ligands are plotted as light orange triangles. CR-ESK1 are plotted as circles and CR-Pr20 as squares. Peptides which were validated by an orthogonal method as bonafide ESK1 ligands are colored red. Peptides which were tested and found not to be ESK1 ligands are colored in dark blue.

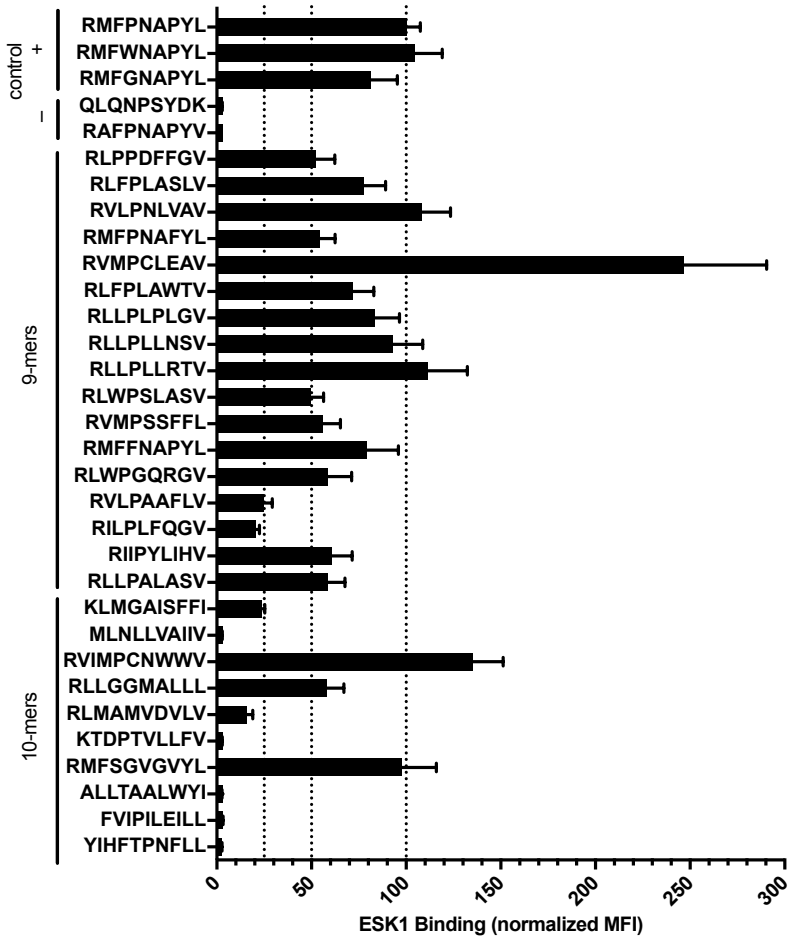


Figure 20 Validation of ESK1 hits by peptide pulsing. 27 peptides that were highly enriched for ESK1 binding and had high predicted affinity to HLA-A*02:01 were synthesized at microgram scale, pulsed onto T2 cells at 50µg/ml (~500µM) and stained with a fluorescently labeled ESK1. Previously identified cross-reactive targets were included as positive controls. The median fluorescence intensity (MFI) of ESK1 binding is plotted, normalized to RMFPNAPYL, set at 100 units.

Recently, large databases of HLA-A*02:01 peptide ligands isolated from tumors and normal tissue have become available[4,41,42,57]. Within these databases (including personal correspondence with Department of Immunology members at Tübingen) we found 48 nine-mer peptides with Arg1 and Pro4. Forty-five had been included in the library of CR-ESK1 peptides identified *in silico* from the human proteome. We synthesized 27 of these and found that 17 (63%) bound to ESK1 when pulsed on T2

cells. The median ESK1 enrichment in the flow-based screen of these 17 peptides was 1.8 whereas the median enrichment of the non-binders was 0.86 **(Figure 21)**, indicating that these ESK1 ligands were enriched in the screen, even if they were not among the top hits.

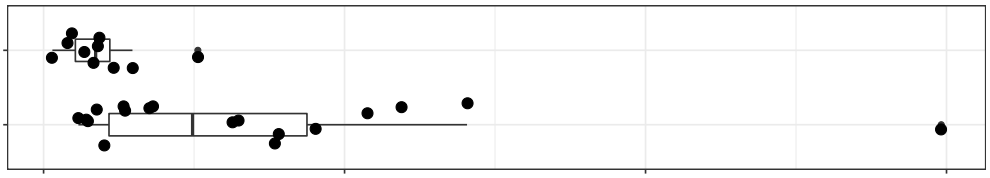


Figure 21 ESK1 screen enrichment scores of bonafide HLA-A*02:01 ligands also found in the set of CR-ESK1 peptides. An additional 27 CR-ESK1 peptides were identified as ligands of HLA-A*02:01 by immunoprecipitation of HLA followed by mass spectrometry to identify peptides. These peptides were synthesized and tested for binding to ESK1. The flow-based ESK1 enrichment scores of the 17 ESK1 binders and 10 non-binders are plotted. p=0.00834 by Welch two sample t-test.

Two WT1-negative[58] cell lines with HLA-A*02:01 ligandomes that have been characterized by immunopeptidomics were found to present ESK1 off-targets discovered in the PresentER screen (TPC-1 cell line: RLPPPFPGL, RVMPSSFFL, RLGPVPPGL, JY cell line: KLYNPENIYL, RLVPFLVEL). RMFPNAPYL was not found among the MHC-I ligands immunoprecipitated from these lines. We tested ESK1 binding in these lines and found that JY cells bound ESK1 at high levels while TPC-1 was marginally positive for ESK1 binding (**Figure 22**).

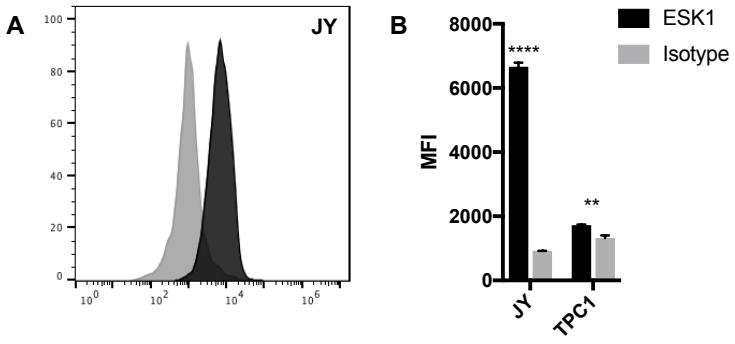


Figure 22 ESK1 binds to two cell lines presenting ESK1 cross-reactive ligands. **(A)** Representative ESK1 and isotype staining of the JY cell line. **(B)** Quantification of ESK1 and isotype staining of the JY and TPC1 cell lines.

A PresentER screen of Pr20 reveals dozens of cross-reactive pMHC

A screen of Pr20 cross-reactive ligands was performed in the same manner as described above. The abundance of each minigene in the unsorted samples was highly correlated (r=0.90-0.94) **(Figure 17 B-D)**. Unlike in the ESK1 screen, positive control peptides that were known Pr20 ligands were not enriched relative to the negative controls (p=0.71) (**Figure 23**). However, twenty peptides were enriched for Pr20 binding at least 5-fold and had predicted ic50s of less than 100nM, suggesting they might be Pr20 off-targets **(Figure 24)**. 45 peptides were synthesized, including the 20 enriched in the Pr20 off-target screen and several additional bonafide HLA-A*02:01 ligands identified by mass spectrometry that matched the Pr20 ligand consensus sequence. Of these, 28 (62%) were found to be Pr20 ligands when the peptides were pulsed onto T2 cells, including 3 known HLA-A*02:01 ligands. The 28 validated peptides were more enriched for Pr20 binding in the screen than the 17 non-validated peptides (**Figure 25**).

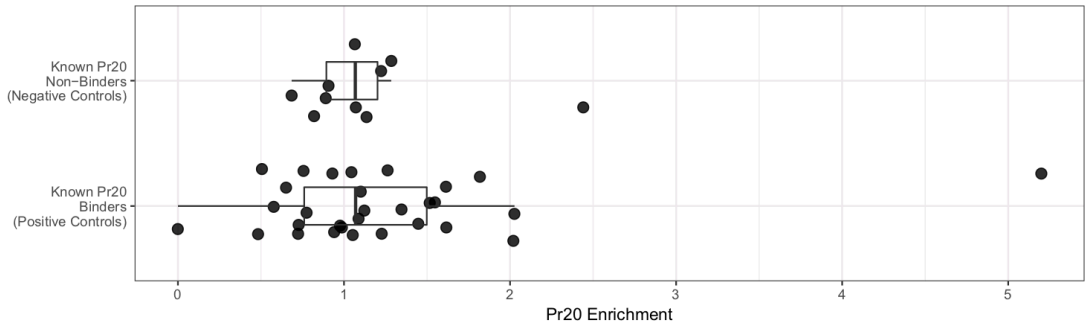


Figure 23 Pr20 screen enrichment scores of positive and negative control Pr20 ligands. The flow screen enrichment scores of control peptides known to be Pr20 binders or non-binders before screening shows no enrichment of positive control peptides compared to negative control peptides. P-value = 0.71 by Welch two sample t-test.

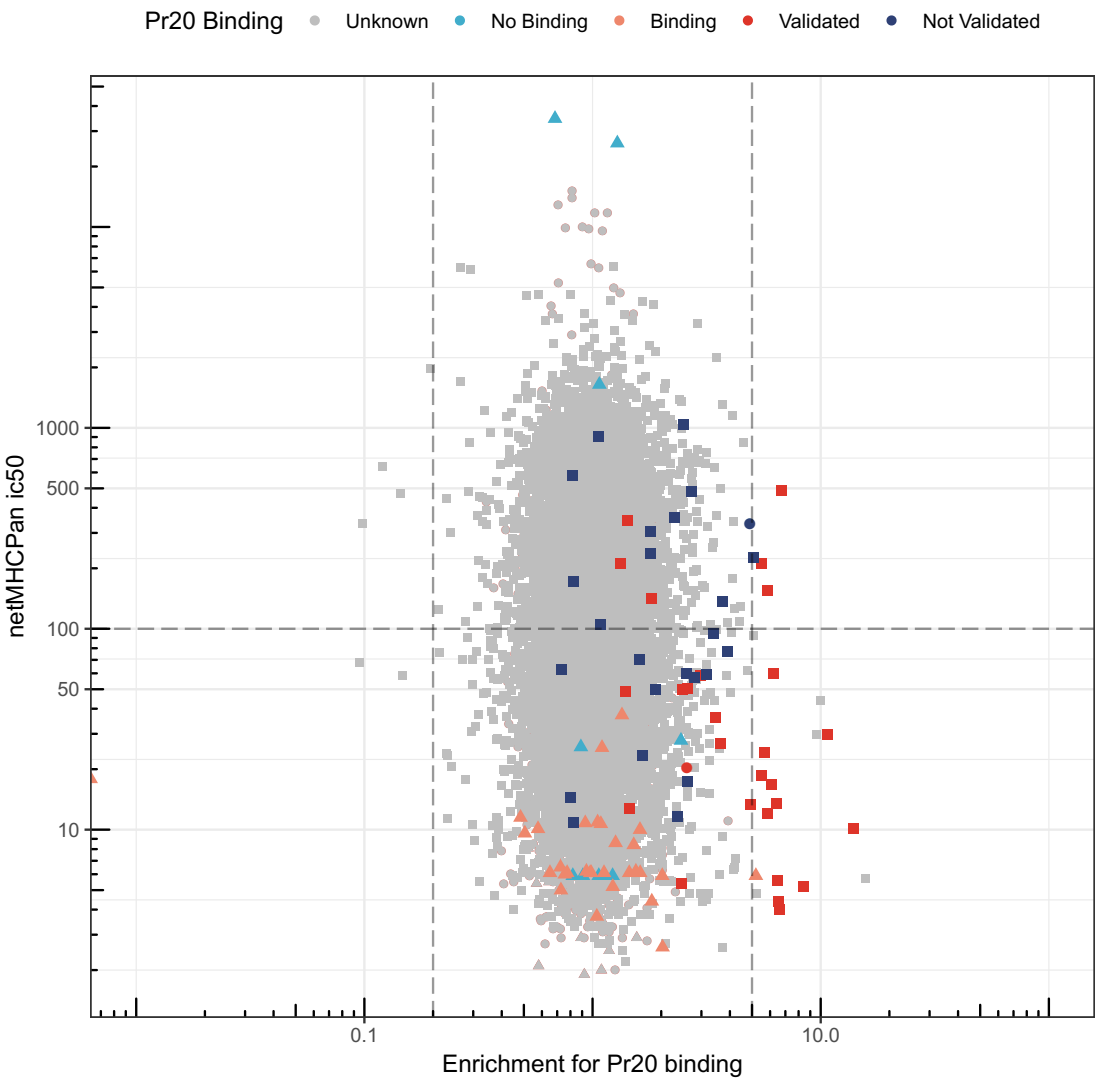


Figure 24 Minigenes enriched in Pr20 screening validate as Pr20 ligands. The netMHCPan predicted HLA-A*02:01 affinity in ic50 (nM) and Pr20 enrichment score of all screened peptides. Each point is a unique peptide minigene. The x-axis indicates enrichment for Pr20 binding (with 1 set as no enrichment) and y-axis indicates the peptide's predicted affinity to HLA-A*02:01 (ic50 nM). Lower ic50 indicates higher affinity. Previously known Pr20 ligands are plotted as light orange triangles. CR-ESK1 are plotted as circles and CR-Pr20 as squares. Peptides which were validated by an orthogonal method as Pr20 ligands are colored red. Peptides which were tested and found not to be ESK1 ligands are colored in dark blue.

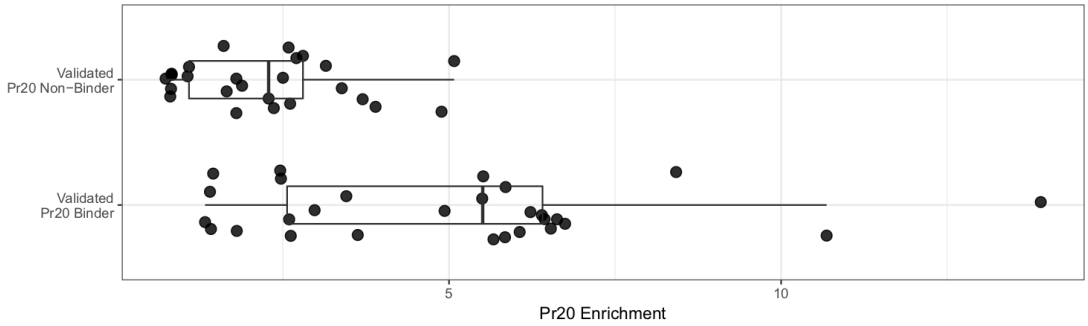


Figure 25 Pr20 screen enrichment scores of peptides that were validated as Pr20 ligands by peptide pulsing. The flow cytometry screen enrichment scores of peptides that were tested for Pr20 binding. Peptides were chosen because of their enrichment in the Pr20 screen or because the peptides were identified in a database of HLA-A*02:01 ligands.

The ESK1 and Pr20 off targets discovered by screening bind more strongly to ESK1 and Pr20 than the intended WT1 and PRAME peptide targets

Examining only the CR-ESK1 subset of peptides, we noticed that the peptides most enriched for ESK1 binding were also predicted to have the highest affinity for HLA-A*02:01 **(Figure 26)**. The peptides that are ≥5-fold enriched for ESK1 binding have a median affinity to HLA-A*02:01 of 31nM, compared to 95nM for the library as a whole and 102nM for ≥5-fold depleted peptides (**Figure 27A**). We found the same result in the Pr20 screen: the most enriched Pr20 ligands also had the highest affinity to MHC-I (**Figure 27B**). We cloned minigenes for four of the most enriched CR-ESK1 (RLFPLAWTV 31.8x; KLMGAISFFI 41.9x) and CR-Pr20 (WLLGDSSFFL 6.5x; LLIQEGPFFV 6.6x) peptides and tested them for binding to ESK1 and Pr20. Compared to the original ESK1 and Pr20 targets—RMFPNAPYL and ALYVDSLFFL— cells expressing these four off-targets bound ESK1 and Pr20 at much higher levels (**Figure 28**).

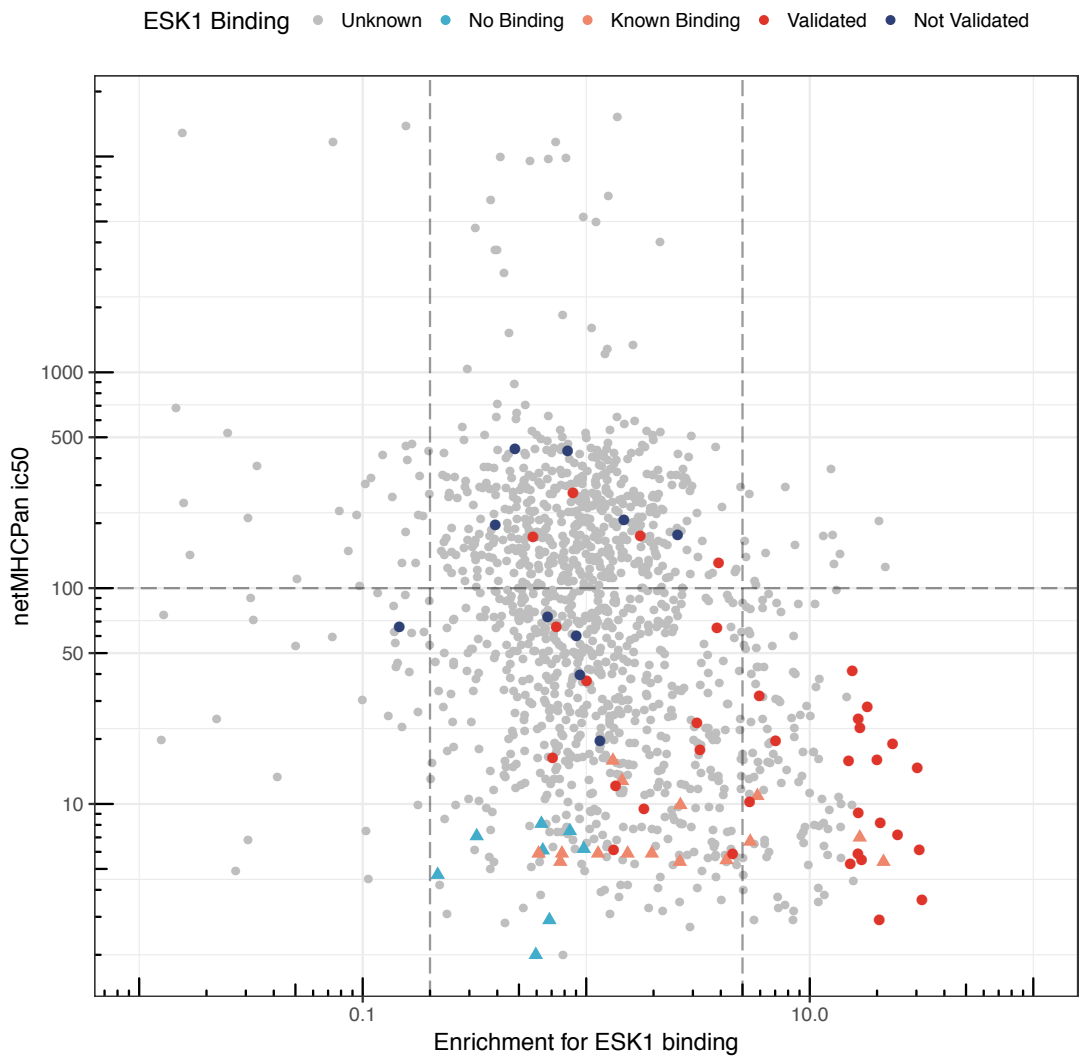


Figure 26 ESK1 screen enrichment scores for CR-ESK1 peptides only. Scatterplot of the ESK1 library screen with only CR-ESK1 peptides (and controls) plotted. Each point is a unique peptide minigene with the x-axis indicating minigene enrichment for ESK1 binding (with 1 set as no enrichment) and y-axis indicating the peptide's predicted ic50 (in nM) to HLA-A*02:01. Lower ic50 indicates higher affinity. Marked control peptides and known ESK1 target peptides are plotted as triangles and CR-ESK1 peptides as circles. Peptides that were validated by an orthogonal method as bonafide ESK1 ligands are colored red. Peptides which were tested and found not to be ESK1 ligands are colored in dark blue.

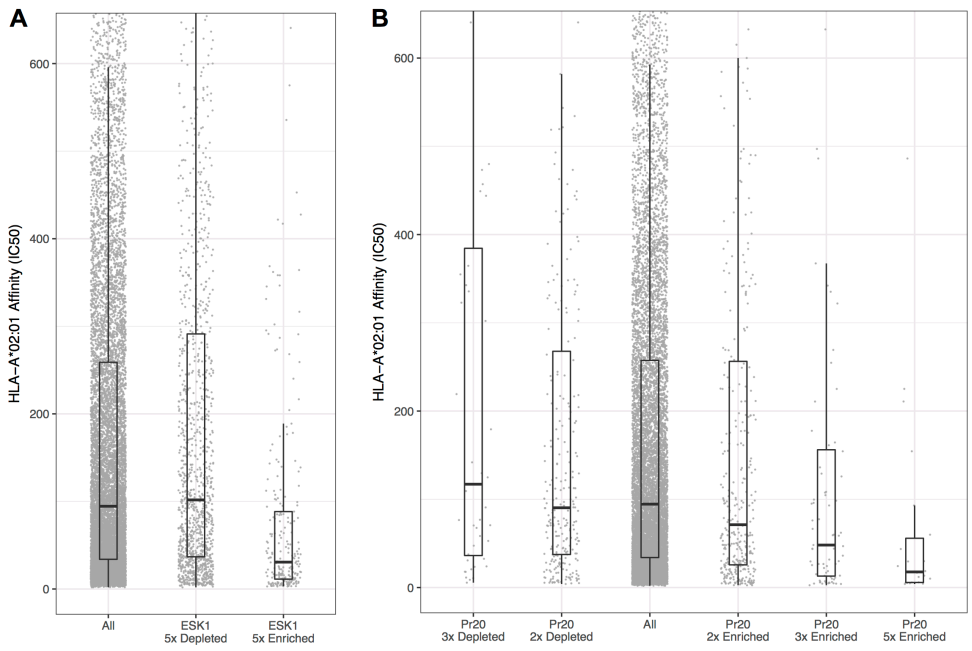


Figure 27 Peptides enriched in TCRm screening are predicted to be high affinity MHC-I ligands. The netMHCPan predicted HLA-A*02:01 affinity in ic50 (nM) of all screened peptides compared to peptides which were ≥5-fold depleted in the ESK screen and peptides that were ≥5-fold enriched for ESK1 binding. **(B)** The netMHCPan predicted HLA-A*02:01 affinity in ic50 (nM) of all screened peptides compared to peptides that were ≥5-fold, ≥3-fold or ≥2-fold enriched in the Pr20 screen and peptides that were ≤3 or ≤2-fold depleted for Pr20 binding.

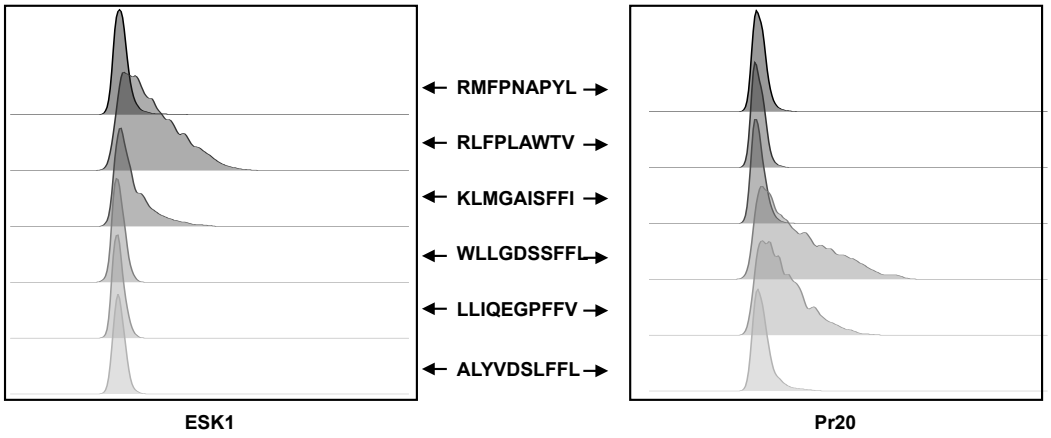


Figure 28 ESK1 and Pr20 staining of outstanding ESK1 and Pr20 ligands.
Four minigenes encoding peptides that were highly enriched in the screen of ESK1 and Pr20 off targets were cloned. T2 cells transduced with the peptide minigenes were stained with fluorescently labeled ESK1 and Pr20 and showed that the minigene encoded antigens were exceptional binders to ESK1 and Pr20. In fact, these antigens bind better to ESK1 and Pr20 than the targets against which these TCR mimic antibodies were designed.

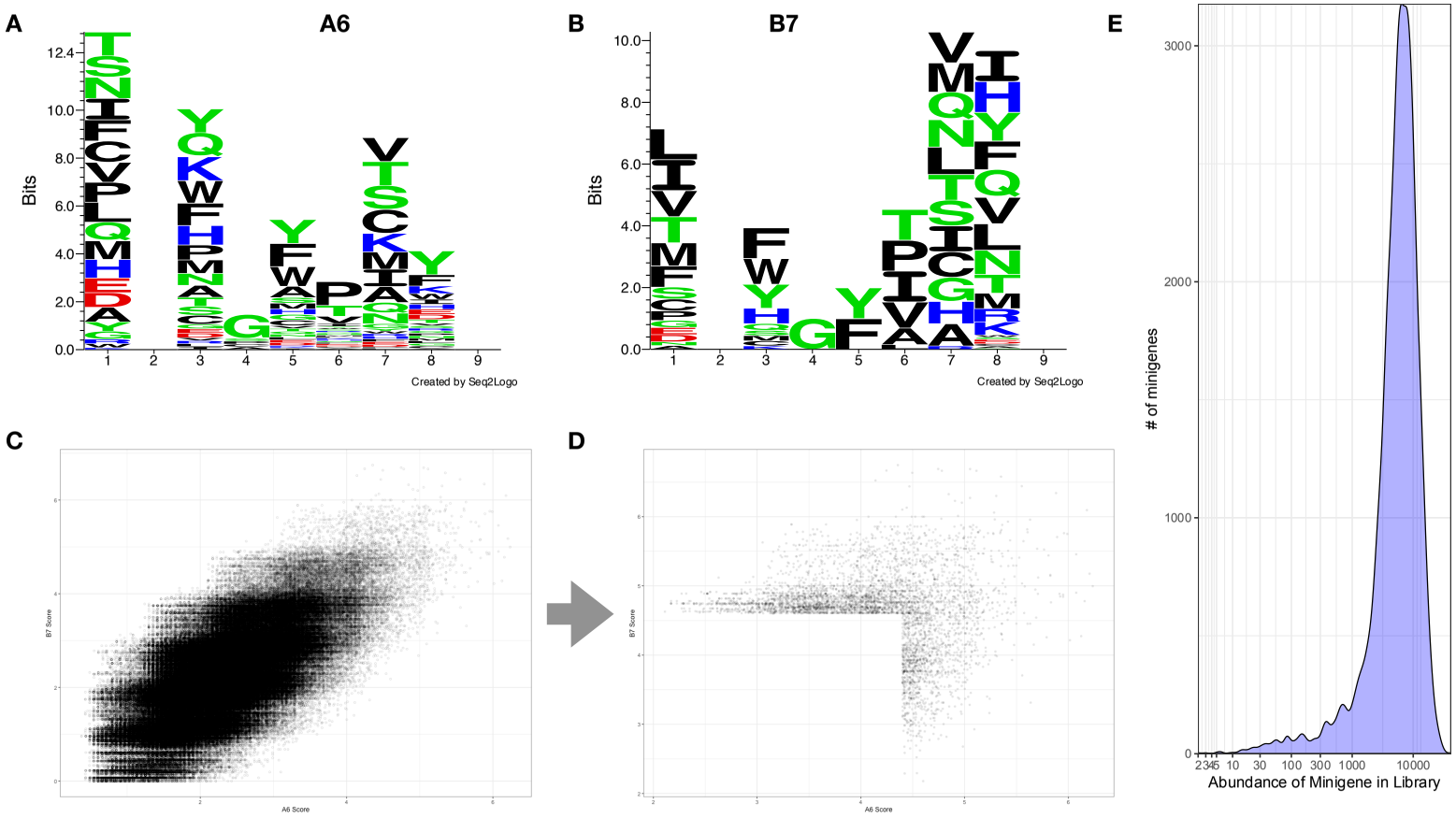
PresentER can be used to discover the targets of T cell receptors

In the previous sections, we have described how pooled screening with the PresentER system may be used to identify both theoretical and, in some cases, actually presented MHC-I peptide off-targets of TCR mimic antibodies. Next, we turned our attention to T cells and asked if we could identify the targets and off-targets of T cells. Human T-lymphotropic virus 1 (HTLV-1) associated myelopathy/tropical spastic paraparesis (HAM/TSP) is an immune-mediated disease primarily of the spinal cord that occurs in a small fraction of individuals infected with the HTLV-1. Patients with HAM/TSP can develop antibodies that recognize both the HTLV-1 Tax protein and the C-terminus of human heterogeneous nuclear ribonucleoprotein A1[59], which is thought to possibly underlie the pathogenesis of the

disease. However, patients also have circulating T cells that recognize an antigenic N-terminal epitope of the virus: Tax 11-19 LLFGYPVYV presented on HLA-A*02:01. It is unknown if there is cross-reactivity between the HTLV-1 Tax-directed TCRs and wild-type peptides presented on HLA-A*02:01 that might be associated with the pathogenesis of HTLV-1. The A6 and B7 TCRs were both derived from a HAM/TSP patient and both recognize the Tax 11-19 peptide, albeit with two different binding modes[60]. These two TCR clones have been well characterized over the past two decades (the A6 TCR is the first human TCR structure to be solved in complex with peptide and MHC-I[7]) and dozens of off-targets that differ from the original peptide by only a single amino acid are known[51,60,61]. We cloned the A6 (also known as "2G4") and B7 TCRs into a mammalian expression vector and transduced donor T cells with these constructs or with the irrelevant 1G4 and DMF5 TCRs. The transduced T cells were functional and could recognize PresentER encoded target cells as assessed by ELISpot (**Figure 9**) and co-culture killing assays (**Figure 10**).

We constructed a library of possible off-targets of A6 and B7 by scoring all 9-mer peptides with NetMHCPan predicted HLA-A*02:01 ic50 of <500nM in the human proteome for likelihood of being targets of A6 and B7, based on preexisting peptide-specific cytotoxicity data (**Figure 29A-C**)[60]. The 5,000 peptides with top A6 or B7 binding scores were selected to make a PresentER minigene library (**Figure 29D**). All single amino acid substitutions to Tax 11-19 LLFGYPVYV were included in the library, as well as several control peptides. The library was constructed as before, sequence validated by Illumina sequencing and is referred to as the "A6B7 library" **(Figure 29E)**. As before, T2 cells were spinoculated at 1/3 MOI with the A6B7 library and transductants were selected with puromycin.

Figure 29 Design of the A6B7 off target PresentER library. The cytotoxicity scores reported in Hausmann et al (1999) for each amino acid substitution to Tax 11-19 LLFGYPVYV were used to construct a sequence logo indicating the consensus **(A)** A6 and **(B)** B7 peptide target. **(C)** All human 9-mer peptides predicted to bind to HLA-A*02:01 (ic50<500nM) were ranked using the A6 (x-axis) and B7 (y-axis) motifs and plotted. **(D)** The 5,000 peptides selected for the A6B7 off-target library are the top predicted A6 and B7 peptides. **(E)** Representative sequencing of the library to confirm normal distribution of the minigenes and no outliers.


A
A6
Bits
Created by Seq2Logo
B
B7
Bits
Created by Seq2Logo
C
B7 Score
A6 Score
D
B7 Score
A6 Score
E
# of minigenes
Abundance of Minigene in Library

A6, 1G4 and DMF5 expressing T cells were co-cultured with T2 cells expressing the A6B7 library for 96 hours in 96 well U-bottom plates. All cells were pooled and re-dispersed across the 96 well plates daily to prevent well-specific outgrowth or depletion. After 96 hours, cells were harvested, genomic DNA was extracted and minigenes sequenced. When we compared T2s that had not been co-cultured with T cells (in triplicate) to T2s that had been cultured with 1G4 or DMF5, we saw excellent correspondence between the samples and no minigenes that were robustly depleted (**Figure 30A-B**). By contrast, over 80 minigenes were depleted at least 2-fold in the library co-cultured with A6 T cells **(Figure 30C)**. Strikingly, all of these depleted peptides were single-residue substitutions of the original Tax 11-19 peptide. For instance, the most strongly depleted peptides were L1S (46x), Y9C (35x), L2C (30x). In general, there was excellent overlap between the peptides discovered by PresentER and the peptides that fit the motif derived from Hausmann as A6 ligands[60] (**Figure 31**).

Figure 30 Abundance of minigenes in A6B7 library after co-culture with DMF5, 1G4 or A6 TCRs. T cells expressing **(A)** DMF5, **(B)** 1G4 or **(C)** A6 were co-cultured with the PresentER A6B7 library for 96 hours and all cells were harvested and minigenes sequenced. Each point is a minigene. Circles are peptides from the human proteome, squares are single amino acid substitutions of the Tax 11-19 peptide and triangles are control peptides. The x-axis indicates the abundance of each minigene in the “no T cells” group while the y-axis indicates the abundance of each minigene after depletion with one of the recombinant T cells.

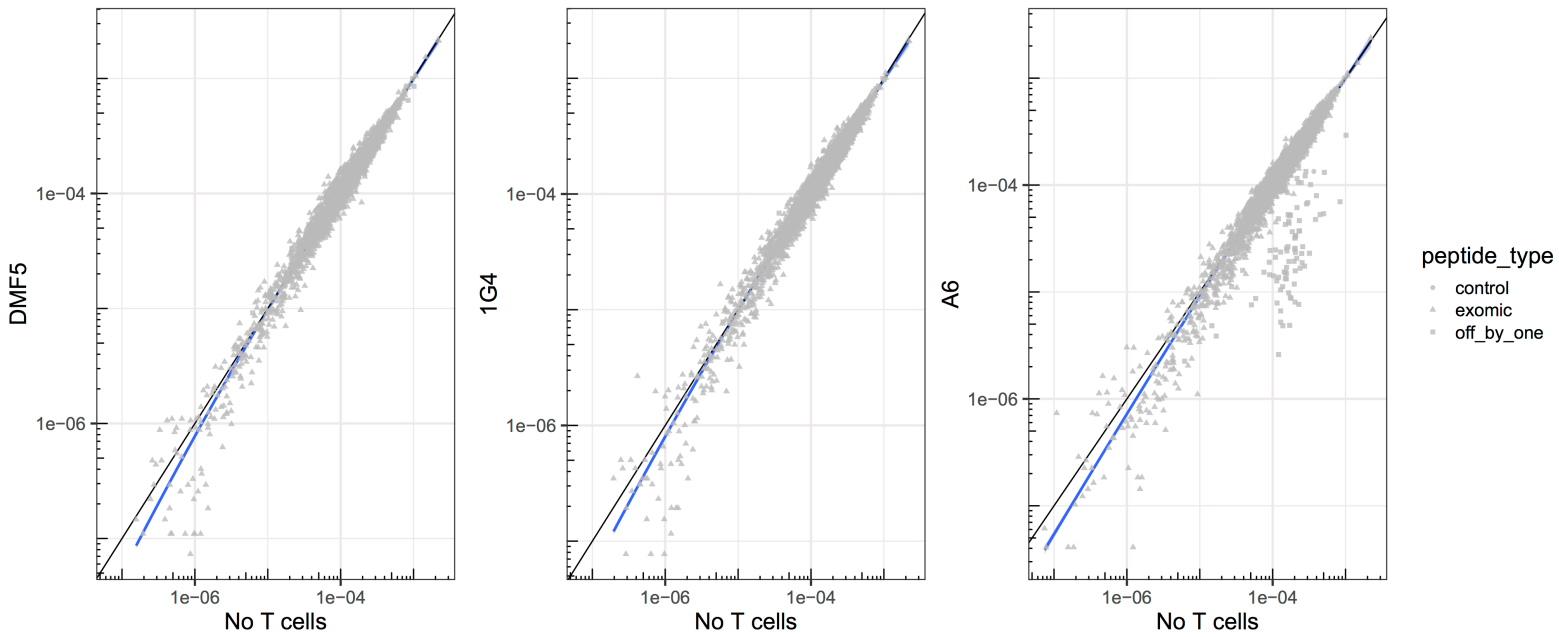
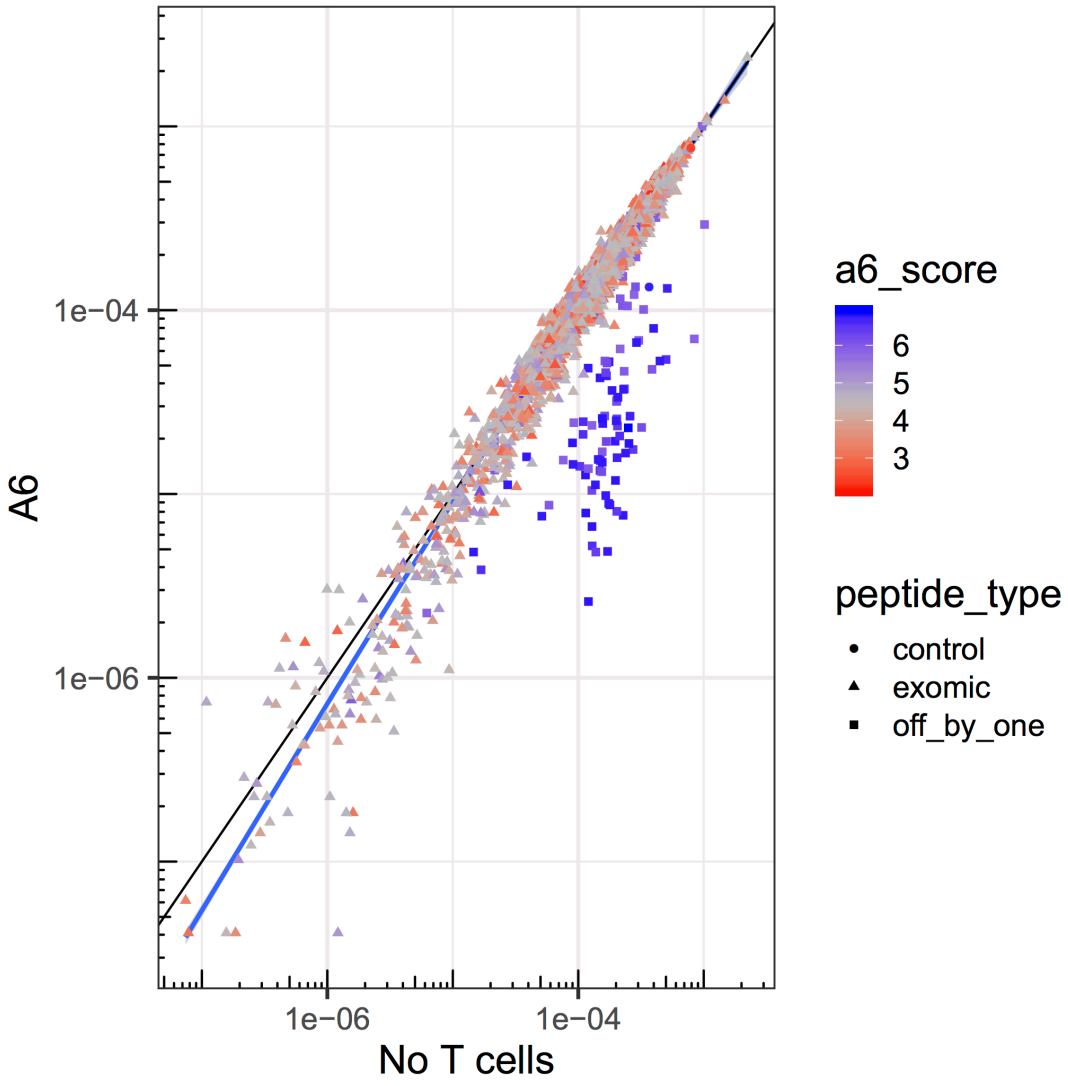
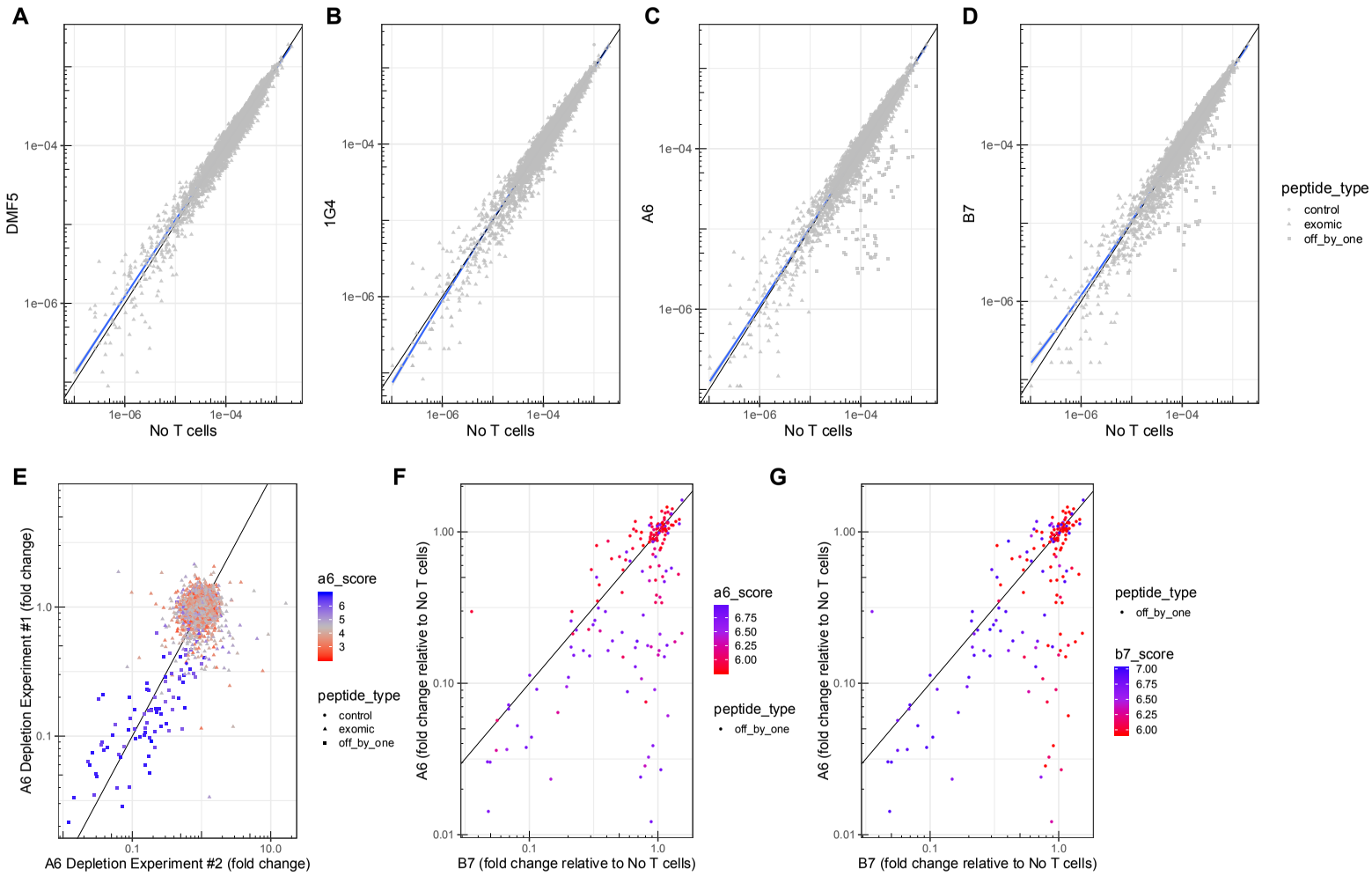

DMF5
1G4
A6
1e−04
1e−06
1e−06
1e−04
No T cells
peptide_type
control
exomic
off_by_one

Figure 31 Abundance of minigenes in A6B7 library after co-culture with A6 library and shaded by A6 score. T cells expressing A6 were co-cultured with the PresentER A6B7 library for 96 hours and all cells were harvested and minigenes sequenced. Each point is a minigene. Circles are peptides from the human proteome, squares are single amino acid substitutions of the Tax 11-19 peptide and triangles are control peptides. All minigenes are shaded according to their similarity to the A6 ligands described by Hausmann[60]


A6
1e−04
1e−06
1e−06
1e−04
No T cells
a6_score
6
5
4
3
peptide_type
control
exomic
off_by_one

The A6B7 library depletion was repeated with freshly transduced PBMCs from a different donor expressing A6, B7, 1G4 and DMF5 TCRs. As in the first co-culture depletion experiment, 1G4 and DMF5 T cells did not deplete any minigenes in the library (**Figure 32A-B**), but both A6 and B7 T cells depleted some single amino acid variants of the Tax peptide (**Figure 32C-D**). There was striking consistency between the peptides depleted by T cells expressing the A6 TCR in the first and second experiments (**Figure 32E**), suggesting that inter-experimental variability of minigene depletion was low. With the inclusion of the B7 TCR in the depletion study, we were able to compare the targets of A6 and B7. 70 and 42 peptides were >2-fold depleted by the A6 and B7 TCRs, respectively, with 36 overlapping peptides. These results are consistent with the more stringent binding requirements of the B7 TCR . Indeed, by color-coding the minigenes according to A6 and B7 binding motifs derived from Hausmann 1999, we can clearly see that the majority of A6 depleted peptides fit the A6 binding motif (**Figure 32E-F**). Similarly, the peptides depleted by both A6 and B7 form a distinct group with the highest possible B7 scores (**Figure 32E-G**). Some peptides with high A6 and B7 scores are not depleted by A6 or B7 T cells, indicating that while they may be ligands when pulsed as soluble peptide onto cells, they are likely not ligands when expressed endogenously.

Figure 32 Abundance of minigenes in A6B7 library after co-culture with DMF5, 1G4, A6 or B7 TCRs. T cells expressing **(A)** DMF5, **(B)** 1G4, **(C)** A6 or **(D)** B7 were co-cultured with the PresentER A6B7 library for 96 hours and all cells were harvested and minigenes sequenced. Each point is a minigene. Triangles are peptides from the human proteome, squares are single amino acid substitutions of the Tax 11-19 peptide and circles are control peptides. The x-axis indicates the abundance of each minigene in the "no T cells" group while the y-axis indicates the abundance of each minigene after depletion with one of the recombinant T cells. **(E)** Comparison of minigenes depleted (relative to the no T cells sample) in the first A6 co-culture experiment and the second A6 co-culture experiment shows excellent correspondence, indicating that the library depletion is robust to inter-experimental variability. **(F-G)** Comparison of the A6 and B7 depletion of the off-by-one peptides shows that most peptides recognized by B7 are also recognized by A6. However, there is a subset of peptides that is only recognized by A6. The peptides are colored by A6 score (**F)** or B7 score **(G)** to show that both scoring systems are predictive of which peptides are recognized by A6 and B7.


A
B
C
D
DMF5
1G4
A6
B7
1e−04
1e−06
No T cells
peptide_type
control
exomic
off_by_one
E
F
G
A6 Depletion Experiment #1 (fold change)
A6 Depletion Experiment #2 (fold change)
a6_score
6
5
4
3
A6 (fold change relative to No T cells)
B7 (fold change relative to No T cells)
6.75
6.50
6.25
6.00
b7_score
7.00
6.75
6.50
6.25
6.00
1.00
0.10
0.01
0.1
1.0
10.0

Discussion

The ability to rapidly identify the targets of T cells and, by extension, TCR like molecules, is a major goal in the field of immunology. Here, we have described a method we term “PresentER” to identify the targets of TCR like therapeutic molecules as well as T cells. The PresentER system generates low levels of antigen presentation relative to peptide pulsing, but sufficient antigen is generated to be detected by fluorescently labeled TCR mimic antibodies and antigen-specific T cells. We screened two TCR mimic antibodies and two engineered T cells against libraries of exomic peptides and found, for the antibodies, hundreds of cross-reactive targets. Although we did not find any human exome-derived peptides that were cross-reactive with T cells expressing the A6 or B7 TCRs, we did re-confirm that these TCRs bind and kill cells expressing 70-80 different single-amino acid substituents of the Tax peptide.

Previously known ligands of ESK1, Pr20 and A6/B7 were not all identified by flow-based or depletion-based screening of PresentER minigenes. In each case, these ligands were originally identified by pulsing T2 cells or HLA-A*02:01 PBMCs with saturating quantities of each peptide. We speculate that while some of these peptides may be biochemical ligands of the TCR/TCRm, they may not be well presented when expressed genetically, either because of inefficient loading onto MHC-I or negative selection during peptide editing, e.g. by TAPBPR[12]. Indeed, the skew we observed in both ESK1 and Pr20-enriched minigenes towards high-affinity HLA-A*02:01 ligands suggests that genetic expression of peptides selects for presentation of ligands with the highest affinities for HLA-A*02:01. This is be an unexpected feature of PresentER, as affinity to MHC-I is the most important factor in determining if a peptide

is presented on MHC-I (although high expression levels may overcome low affinity[11]). This suggests that methods for generating large number of pMHC *in vitro* by covalently attaching a peptide to the MHC may lead to artificial presentation of peptides that would never be presented endogenously. Additional study of the difference between genetic expression of MHC-I ligands and peptide pulsing should be pursued to help investigators decide which cross-reactive peptides are likely to be endogenous MHC-I ligands and thus pose a risk to patients in a clinical setting.

Although we did not identify any human peptide ligands of the A6 TCR, we did demonstrate the PresentER can be used to identify the targets of T cells in a pooled co-culture screen. We were able to re-confirm 70-80 peptides as targets of the A6 TCR. Interestingly, cells expressing PresentER-MLWGYLQYV (Yeast Tel1p) and PresentER-LGYGFVNYI (Human HuD / ELAVL4) were not depleted in the A6B7 library when co-cultured with A6 T cells, nor were they depleted in individual co-culture assays. Moreover, several other substituted peptides that had been reported by Hausmann to bind A6 did not deplete in the library assay. For instance, the Tel1p and HuD/ELAVL4 peptides were identified by Hausmann as cross-reactive targets of A6 by peptide pulsing. However, biochemical studies of soluble A6 binding to Tel1p and HuD have shown the A6/peptide/HLA-A2 complex to be weakly stable and less favorable than the interaction between Tax/HLA-A2/A6[50,62]. Thus, we hypothesize that the discrepancies between the results of our screen and the findings of Hausmann reflect the much stricter constraints imposed on peptide loading and more physiologic levels of antigen presentation by endogenous peptide presentation in contrast to exogenous peptide pulsing

Preclinical evaluation methods for novel therapeutic agents directed towards peptide-MHC have been insufficient to prevent harmful off-tumor off-target toxicity. Indeed, even recent reports of engineered anti-cancer TCRs show preclinical evaluation of only hundreds of off-target peptides, most of which are merely single amino acid substitutions[63,64]. In PresentER, we have developed a mammalian screening approach to prospectively identify thousands of cross-reactive MHC-I ligands. While PresentER may not detect all naturally expressed cross-reactive epitopes and may detect epitopes that are never presented in an endogenous setting, the system can help to identify potential cells and tissues at risk for closer clinical surveillance. The PresentER system is easy to scale-up and a sufficiently well-funded company could survey the entire human proteome for cross-reactive ligands. Large libraries of PresentER encoded ligands spanning the whole human proteome could be used to discover the off-targets of TCR agents before any patients is exposed to toxicity. Using the relatively non-conservative 500nM cut-off for HLA-A*02:01 binding, there are ~750,000 possible HLA-A*02:01 peptides that can be found in the human proteome (**Figure 5**). 75 libraries organized in conveniently assayable sets of 10,000 peptides could be synthesized for less than $50,000 and an engineered/cloned TCR could be tested against this library for $100,000-$200,000 of reagent and sequencing costs. This cost is a fraction of the billions of dollars typically spent during drug development[65] and would go far to ensure that such drugs are safe to give to the first patient on a phase I clinical trial.

We have shown that PresentER can be used for biochemical evaluation of potentially therapeutic TCR based agents. PresentER can also be used as an immune presentation platform *in vitro* and *in vivo*; thus, this work can be expanded to recapitulate the MHC restricted antigenic diversity of human cancer. Libraries of

MHC-I ligands could be used to ask how tumor neoantigen heterogeneity affects progression and treatment of tumors in immunocompetent animals, and address areas such as neoantigen immunogenicity and clonality, cancer vaccination and immunoediting.

Chapter II: Rejection of immunogenic tumor clones is limited by clonal fraction

Introduction

Human cancers bear uniquely distinguishable features on the surface of their cells in the form of neoantigens, comprised of peptides derived from mutated, foreign or oncofetal proteins that are presented in complex with Major histocompatibility complex I (MHC-I) molecules. These short, 8-11 amino acid fragments can specifically mark cancer cells and activate potent immune responses that, when harnessed, can lead to effective anti-cancer therapy[19,21,66]. Karl Hellström first described the coexistence of tumor-specific lymphocytes together with cancer cells in human solid tumors as a paradox 50 years ago[67]. In mice, an analogously enigmatic observation has been made that sporadic tumors occurring in aged or carcinogen-treated mice induce strong T cell responses when transferred into new hosts, preventing engraftment[68-70]. Although it is known that tumor-specific T cells can lead to tumor rejection in a new host or when used as a cancer therapy, it is not understood how and to what extent immune surveillance is evaded during early tumorigenesis in the original host that developed an immunogenic tumor. The increased rate of tumor formation in immunocompromised individuals has led to the hypothesis that the immune system can and does eliminate some tumors, particularly virally induced tumors, before they become clinically apparent[71]. We hypothesize that if the immune system can eliminate some early tumors, but not others, perhaps it is because some antigens are more potent at inducing effective T cell responses during early tumorigenesis. Identification and characterization of neoantigens that can induce an effective immune response and clear cancer cells is critical to

understanding why and how immunogenic tumors develop in immunocompetent hosts.

Some immunogenic peptides have been discovered in animal studies by injection of thousands or millions of tumor cells bearing neoantigens into animals and observing tumor rejection. However, the robust immune activation and tumor rejection in these cases is not analogous to the events of early tumorigenesis in humans, when the number of transformed cells is miniscule. Thus, even though some tumors are immunogenic, it is not clear why the host cannot eliminate them when the tumors first arise. If the biochemical features of neoantigens that lead to effective T cell responses were known, it might be possible to identify which tumors bear immunogenic antigens. Some reports have linked peptide immunogenicity to the characteristics of some residues at certain positions along the MHC-I ligand[72], while others have focused on the difference between the affinity of a wild-type ligand and a mutated ligand[73]. However, the absence of a large, unbiased data set of known immunogenic and non-immunogenic peptides has stymied the validation of these approaches. Indeed, most known immunogenic antigens are derived from viral proteins and few mutationally-derived neoantigens are confirmed as bonafide immunogenic peptides in mice or humans. Here, we have used the PresentER method to express libraries of precisely defined MHC-I ligands in mammalian tumor cells and have used this method to ask questions about MHC peptide immunogenicity in immunocompetent animals during early tumorigenesis.

Using libraries of PresentER encoded MHC-I ligands, we tracked the dynamic growth and depletion of thousands of tumor subclones *in vivo* and noted a striking failure of cancer immunosurveillance that is potentially analogous to the failure of immune

surveillance in humans during early tumorigenesis. We demonstrate for the first time that the ability of the naive immune system to surveille a nascent tumor and reject immunogenic subclones is limited by the fraction of cells expressing each unique antigen. Furthermore, we show that these rejection thresholds vary among antigens. Our data are consistent with the observation in humans that patients whose tumors have high numbers of subclones—and thus more subclonal neoantigens—have increased levels of relapse and worse survival than patients with more homogenous tumors[74-76]. Thus, our data provide an antigen-specific rationale for the impact that tumor heterogeneity has on survival of human patients. According to our findings, antigen-specific immune effects are limited during early tumorigenesis, which has implications for the emergence and outgrowth of immunogenic tumors.

Materials and Methods

Animal Studies

6-8 week old C57BL6/N mice were purchased from Envigo or Taconic Biosciences. 6-8 week old B6.SJL-*Ptprc*^{a}/BoyAiTac (known as CD45.1 mice) and B6.129S6-Rag2^{tm1Fwa} N12 Mice (known as RAG2 KO) were purchased from Taconic Biosciences. Mice were shaved before subcutaneous engraftment of indicated number of RMA/S cells in 100 uL PBS. Tumor volumes were calculated using caliper measurements and the standard modified ellipsoid formula: tumor volume = (LxW^2)x0.52 every 2-3 days. Animals were euthanized when tumor volume exceeded 2000 mm^3 or if ulceration was noted. Vaccination of animals was performed by subcutaneous injection of $10x10^6$ irradiated (20 Gy) MCA205-ΔTAP2 cells expressing libraries of minigenes.

Flow Cytometry

For cell surface staining, cells were incubated with appropriate fluorophore-conjugated mAbs for 30 minutes on ice and washed twice before resuspension in the viability dye DAPI at 1 μg/mL. Flow cytometry data were collected on a LSRfortessa (BD) or an Accuri C6 (BD) and analyzed with FlowJo V10 software. The antibodies used in this study were anti-H-2Kb-APC clone AF6-88.5 (Biolegend 116517), anti-SIINFEKL/H2kb-APC clone 25-D1.16 (Biolegend 141605), anti-CD45.2-APC clone 104 (eBioscience 17-0454-81), anti-CD8a-FITC clone Ly-2 (BD Pharmingen 553030), anti-CD3-PerCP clone 145-2C11 (BD Pharmingen 553067). The following

fluorescently labeled H-2Kb tetramers were obtained through the NIH Tetramer Core Facility: SIINFEKL, SIYRYYGL, MSIIFFLPL, SNFVFAGI and VTFVFAGL.

Generation of MCA205-ΔTAP2

A guide RNA sequence targeting murine TAP2 (ATGGGGCTGTTGCGCTGAGC) was cloned into the LentiCRISPRv2[77] plasmid (Addgene plasmid 52961), a gift from Feng Zhang (Broad Institute, Cambridge, Massachusetts, USA). MCA205 fibrosarcoma cells were transiently transfected using Lipofectamine 2000 (Thermo Fisher Scientific 11668027) following standard manufacturer's protocols. 24 hours later, successful transfectants were selected using 5 ug/mL Puromycin for 3 days before expansion and single-cell subculture. Genetic ablation of TAP2 was verified by next generation sequencing of the TAP2 loci confirming a frameshift deletion in both alleles, and RT-PCR analysis with the following primers (TAP2 For: CACAGCCACCACAAGGAAGA, TAP2 Rev: CAGTTCCTGTCCAGTCGCAT, mGAPDH For: TGATGGGTGTGAACCACGAG, mGAPDH Rev: TCTTGCTCAGTGTCCTTGCT). Reduced cell-surface H-2Kb expression was also verified by flow cytometry.

Determination of antigen-specific immunogenicity thresholds

RMA/S cells expressing PresentER antigen #1 (eGFP) were mixed with RMA/S cells expressing PresentER antigen #2 (mCherry) at defined ratios (e.g. 1:10, 1:100, 1:1000, etc) and validated by flow cytometry immediately before injection into CD45.1 mice. After 17 days, tumors were harvested, cut into small pieces and disaggregated by incubation at 37ºC with Liberase (TL (Sigma-Aldrich 5401020001), DNAse I (Worthington Biochemical LS002139) and ACK lysis buffer (Thermo Fisher Scientific A1049201). Single cell suspensions of tumor cells were stained with CD45.2 and DAPI and collected on the same flow cytometer, using the same settings and gates

as on day 0. The number of eGFP and mCherry cells was calculated based on the same gates used on the day of injection. Non-fluorescent cells were ignored for the purposes of analysis and the percentage of eGFP and mCherry cells in the tumor were normalized to sum to 100%. The fold change in cells expressing each antigen was calculated as: (percentage of normalized eGFP cells in tumor) / (percentage of eGFP cells at D0).

PresentER minigene cloning and transduction of RMA/S

The individual PresentER minigenes specified above were cloned into the PresentER backbone as previously described[78] using the oligonucleotide template 5'-GGCCGTATTGGCCCCGCCACCTGTGAGCGGG…[27-30nt insert]…TAAGGCCAAACAGGCC-3' and the table of PresentER constructs above. HEK293T Phoenix-Ampho cells were transfected with each plasmid and, after 24h, viral supernatant was harvested every 12 hours. RMA/S were transduced with limiting amounts of viral supernatant at 1,000xg for 2h at 37ºC in 6 well non-tissue culture treated plates. PresentER minigenes vectors are available on Addgene (102942, 102943, 102944, 102945, 102946).

Library construction in silico

The peptides included in the libraries were found in Uniprot database of canonical mouse protein sequences (UP000000589). Substrings of unique 8 amino acid sequences were collected and affinity to H-2Kb was calculated using NetMHCPan v4.0. 5,000 randomly selected peptides with predicted ic50<500nM were selected and constitute the "wild type peptide library." A single random amino acid substitution was made to each member of the wild type library to generate the "mutant peptide library." Substitutions which generated another wild type peptide were excluded.

Library generation in vitro

Libraries of PresentER minigenes were cloned as previously described[78]. Library metadata is provided as a supplemental file. Oligonucleotide libraries were ordered from CustomArray and amplified with Phusion polymerase using pool-specific primers: WT library forward 5'-CATGTTGCCCTGAGGCACAG-3' and reverse 5'-CGGATCGTCACGCTAGGTAC-3'. Mutant library forward 5'-GGTCGTCGCATCACAATGCG-3' and reverse 5'-CGGATCGTCACGCTAGGTAC-3'. Library oligonucleotide format: 5'—[pool specific F primer]—GGCCGTATTGGCCCCGCCACCTGTGAGCGGG—[27-30nt insert]—TAAGGCCAAACAGGCC—[pool specific R primer]—3'. Amplicons were digested with SfiI and passed through a MinElute column. The PresentER cassette vector was also digested with SfiI, treated with calf intestinal phosphatase and ligated to the oligonucleotides with T4 ligase. Ligation products were phenol extracted and electroporated into DH5 electrocompetent cells. Electroporated cells were plated and counted. At least 1,000x fold more transformants than minigene library members were required to proceed to plasmid DNA extraction (>$5x10^6$ colonies). The colonies were scraped off the plate and grown for 3.5h in TB + ampicillin at 37ºC at 225rpm. The bacteria were maxiprepped using the Qiagen maxiprep kit and library representation was checked by Illumina sequencing. Library containing retrovirus was produced by transfection of 15cm plates of HEK293T Phoenix-AMPHO cells with library plasmid DNA. Viral supernatants were collected beginning at 24h after transfection and continuing every 12 hours until 72 hours post transfection. Viral supernatants were pooled, concentrated with Clontech Retro-X concentrator and frozen. Concentrated viral supernatant titers were determined by transduction of RMA/S cells. Libraries of minigene expressing RMA/S cells were generated by transduction at an MOI < 0.3

(1,000xg for 2h at 37ºC in 6 well non-tissue culture treated plates) and selection with puromycin. Library expressing cells were maintained in cultures of >1,000x cells per number of minigenes in the library (i.e. at least $5x10^6$).

Genomic DNA extraction and minigene sequencing

To verify that minigene representation was not compromised during cloning, all libraries were sequenced from plasmid prior to transduction into mammalian cells. Minigenes were either amplified with barcoded primers (P5 F: 5' AATGATACGGCGACCACCGAGATCT 3'; P7 Barcoded R: 5' CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCT CTTCCGATC 3' and directly submitted for Illumina HiSeq sequencing using a custom primer (5' ACGCTCTTCCGATCTTTGGCCTGTTTGGCCTTA 3') or they were amplified with a non-barcoded, nested PCR protocol (Primer Set #1: F 5' AATGATACGGCGACCACCGAGATCT 3'; R: 5' GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC 3' and Primer Set #2: F 5' GCCACCTGTGAGCGGG 3'; R: 5' TCTTTGGCCTGTTTGGCCTTA 3') followed by Illumina library preparation and sequencing at the Integrated Genomics Operation at MSKCC. Genomic DNA was extracted from cultured cells or mechanically disaggregated tumors with the Gentra Puregene kit and minigenes were amplified from genomic DNA. Reads were mapped to the PresentER minigene libraries with Bowtie2 using default settings. Reads that did not map to the minigenes in the library were discarded. Quality control to ensure library representation and absence of contamination were performed on every sample. All data analysis was performed in R.

Data Availability

Sequencing data for each of the experiments has been deposited:

Minigene sequencing of RMA/S tumors expressing libraries of PresentER peptide antigen minigenes after 17 days of growth in immunocompetent mice. DOI:10.5281/zenodo.1308910
Minigene sequencing of RMA/S tumors expressing libraries of PresentER peptide antigen minigenes after 17 days of growth in immunocompetent mice that were either vaccinated or nonvaccinated. DOI:10.5281/zenodo.1309837
Minigene sequencing of RMA/S tumors grown in RAG deficient mice, harvested and then transferred and grown in WT or RAG mice. DOI:10.5281/zenodo.1310902

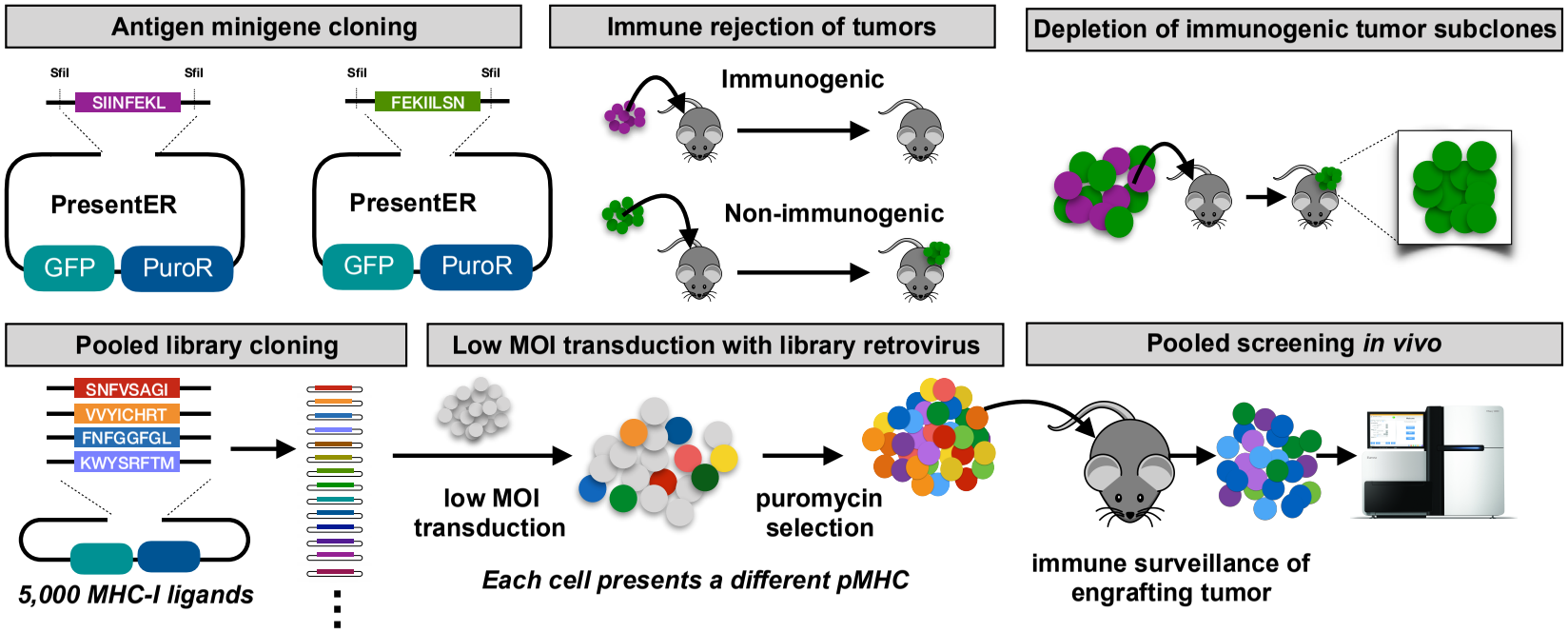
Results

PresentER expressing cells recapitulate known T cell immunogenicity

We have developed a method for encoding diverse peptide/MHC (pMHC) ligands in mammalian cells (termed "PresentER") and used this method to perform high-throughput, pooled screening of MHC-I ligand immunogenicity in wild-type mice. An overview and schematic of the cloning strategy, preliminary findings and approach is presented in **Figure 33**. As described Chapter I, The PresentER antigen minigene is comprised of an endoplasmic reticulum (ER) signal sequence followed by a short peptide/epitope. Expression of the peptide and its display on MHC-I does not require proteasomal degradation or peptide processing, thus enabling precise definition of the exact epitope displayed to the immune system. Transporter associated with antigen presentation (TAP) deficient cell lines expressing PresentER antigen minigenes lead to surface presentation of the encoded MHC-I peptide, detectable by multiple modalities, including fluorescently labeled antibodies directed to specific MHC-I ligands, mass spectrometry based immunopeptidomics and antigen-specific T cell reactivity. To demonstrate the applicability of PresentER antigen minigenes to study MHC-I ligand immunogenicity, we first asked if cancer cells encoding known immunogenic (mouse gp75 TAYRYHLL W223A,H224Y[78]; mouse p68 SNFVFAGI S551F[69]; synthetic SIYRYYGL[79]; synthetic VTFVFAGL[68]; chicken ovalbumin SIINFEKL) or non-immunogenic (mouse PEDF MSIIFFLPL[80]; scrambled chicken ovalbumin FEKIILSN; mouse dEV8 EQYKFYSV[82]; mouse p68 SNFVSAGI[68]; mouse gp75 TWHRYHLL[79]; Mouse Trp2 SVYDFFVWL) MHC-I ligands would be rejected by wild-type (WT) animals. The C57BL/6 syngeneic, TAP deficient mouse cell line

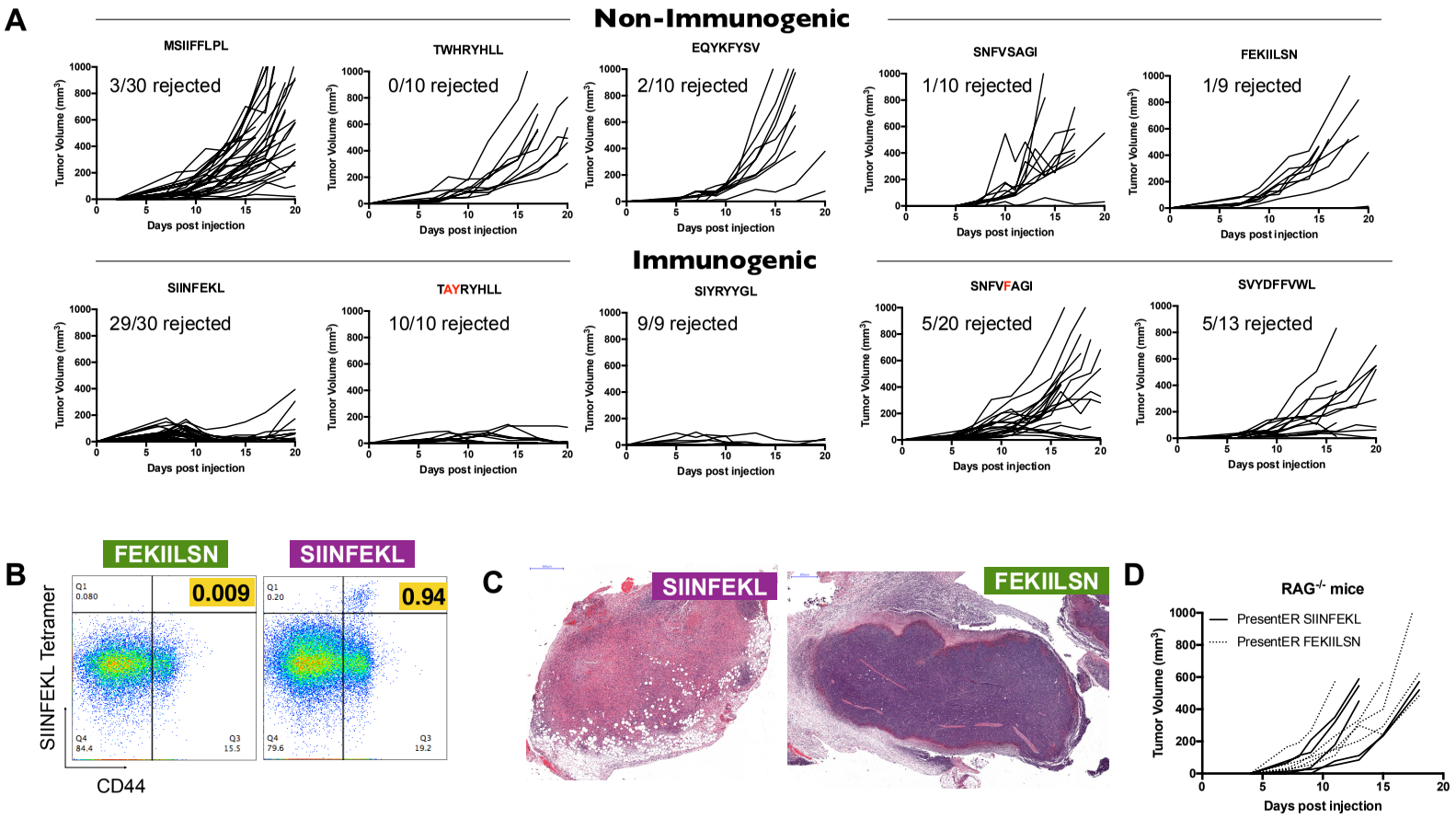
RMA/S was transduced with these antigen minigenes and $5x10^6$ cells were injected subcutaneously into WT C57BL6/N mice. Tumors expressing known non-immunogenic peptides grew, while tumors expressing known immunogenic peptides were rejected in some or most animals (**Figure 34A**). TAP deficient cells were used because peptides derived from proteins that are not directed to the ER (e.g. eGFP) are not presented on MHC as these cells lack the ability of transport peptide from the cytoplasm into the endoplasmic reticulum.

Figure 33 Schematic of the cloning strategy of PresentER antigen minigenes along with the experiments performed *in vivo* in this chapter.


Antigen minigene cloning
SfiI
SfiI
SIINFEKL
PresentER
GFP
PuroR
SfiI
SfiI
FEKIILSN
PresentER
GFP
PuroR
Immune rejection of tumors
Immunogenic
Non-immunogenic
Depletion of immunogenic tumor subclones
Pooled library cloning
SNFVSAGI
VVYICHRT
FNFGGFGL
KWYSRFTM
5,000 MHC-I ligands
Low MOI transduction with library retrovirus
low MOI
transduction
puromycin
selection
Each cell presents a different pMHC
Pooled screening in vivo
immune surveillance of
engrafting tumor

At 7 days after tumor injection, T cells specific for the chicken ovalbumin peptide SIINFEKL could be detected in tumor draining lymph nodes of animals injected with PresentER-SIINFEKL expressing cells (**Figure 34B**). Haemotoxylin and Eosin (H&E) staining of regressing PresentER-SIINFEKL tumors showed lymphocytic infiltration with hyalin rich fibrin deposits, indicating cell death. By contrast, tumors with PresentER-FEKIILSN (a non-MHC-I binding peptide) were well-vascularized, highly cellular and with trace lymphocytic infiltration (**Figure 34C**). In order to verify that the mechanism of tumor rejection was indeed T cell dependent, we injected $RAG^{-/-}$ animals with SIINFEKL or FEKIILSN positive tumors and confirmed that SIINFEKL tumors were not rejected **(Figure 34D)**. Taken together, these results indicate that RMA/S tumors expressing PresentER antigen minigenes can recapitulate the known immunogenicity of individual mouse MHC-I ligands in a T cell dependent manner.

Figure 34 Cells expressing PresentER minigenes recapitulate the known immunogenicity of encoded antigens. (A) $5x10^6$ RMA/S cells expressing a PresentER minigene were injected subcutaneously into C57BL6/N mice and tumor size was monitored by caliper measurements. Top row: WT/non-binding peptide minigenes. Bottom row: mutated/foreign peptide minigenes. All plots are compilations of several experiments. **(B)** Mice were injected with $5x10^6$ RMA/S cells expressing PresentER-SIINFEKL or PresentER-FEKIILSN. Tumor draining lymph nodes were harvested 7 days later and stained with a SIINFEKL/H-2Kb tetramer and CD44. **(C)** H&E staining of tumors expressing PresentER-SIINFEKL or PresentER-FEKIILSN. **(D)** RMA/S PresentER-SIINFEKL and PresentER-FEKIILSN cells were injected into $RAG^{-/-}$ mice and tumor growth was monitored.


A
Non-Immunogenic
MSIIFFLPL
3/30 rejected
TWHRYHLL
0/10 rejected
EQYKFYSV
2/10 rejected
SNFVSAGI
1/10 rejected
FEKIILSN
1/9 rejected
Immunogenic
SIINFEKL
29/30 rejected
TAYRYHLL
10/10 rejected
SIYRYYGL
9/9 rejected
SNFVFAGI
5/20 rejected
SVYDFFVWL
5/13 rejected
Tumor Volume (mm3)
Days post injection
B
FEKIILSN
SIINFEKL
0.009
0.94
SIINFEKL Tetramer
CD44
C
SIINFEKL
FEKIILSN
D
RAG-/- mice
PresentER SIINFEKL
PresentER FEKIILSN
Tumor Volume (mm3)
Days post injection

Tumors expressing PresentER antigens do not cause abscopal rejection, but do cause subclone fraction-dependent bystander killing

To study immunogenicity *in vivo* at high throughput and complexity in this model, we first wanted to understand if immune responses directed at immunogenic antigens lead to rejection of cells presenting non-immunogenic antigens. If so, the ability to study immunogenicity in a pooled *in vivo* setting might be compromised. Mice were injected with pairs of tumors expressing an immunogenic and a non-immunogenic antigen minigenes (one minigene-expressing tumor on each flank): SIINFEKL/FEKIILSN, SIYRYYGL/EQYKFYSV and TAYRYHLL/TWHRYHLL. The immunogenic SIINFEKL, SIYRYYGL and TAYRYHLL expressing tumors were rejected, but the non-immunogenic FEKIILSN, EQYKFYSV and TWHRYHLL tumors found on the contralateral flank were not (**Figure 35**). This suggests that tumor rejection is local and that an effective immune response to an immunogenic tumor does not affect the growth of a non-immunogenic tumor.

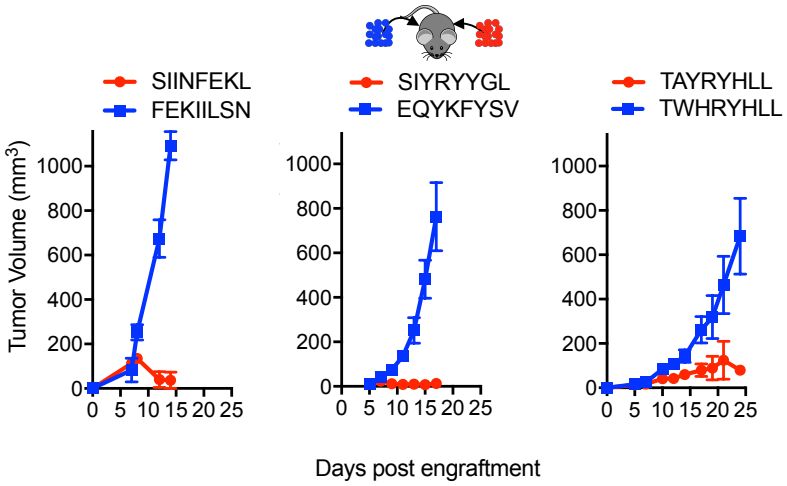


Figure 35 No abscopal effect in the RMA/S antigen minigene tumor model. Mice were injected simultaneously with $5x10^6$ RMA/S cells expressing an immunogenic PresentER minigene on one flank and a non-immunogenic minigene on the contralateral flank: SIINFEKL/FEKIILSN (left; n=3), SIYRYYGL/EQYKFYSV (middle; n=5) and TAYRYHLL/TWHRYHLL (right; n=5). Tumor growth curves are shown.

Next, to test if the immune system could identify and kill immunogenic subclones within a largely non-immunogenic tumor, we generated tumors with varying ratios of immunogenic (SIINFEKL) and non-immunogenic (FEKIILSN) cells. These heterogeneous tumors were injected subcutaneously into mice and tumor size was monitored. In tumors where PresentER-SIINFEKL cells were greater than 25% of a tumor, tumors were smaller and tumor rejection occurred more frequently, especially when tumors were comprised of ≥50% immunogenic cells (**Figure 36A**). Next, we wanted to test if immunogenic subclones within a largely non-immunogenic tumor could be eliminated. We cloned mCherry into the PresentER vector and mixed PresentER-SIINFEKL (mCherry) with non-immunogenic PresentER-MSIIFFLPL

(eGFP) cells at varying ratios and injected them into congenically marked CD45.1 mice. On day 16, the tumors were harvested and flow cytometry was performed to identify which (CD45.2 positive) tumor cells remained. Remarkably, within a non-immunogenic (eGFP labeled) tumor, immunogenic (mCherry labeled) sub-populations were eliminated (**Figure 36B**). Thus, in this model, the immune system is capable of recognizing and selectively depleting immunogenic tumor subclones within the context of a largely non-immunogenic tumor.

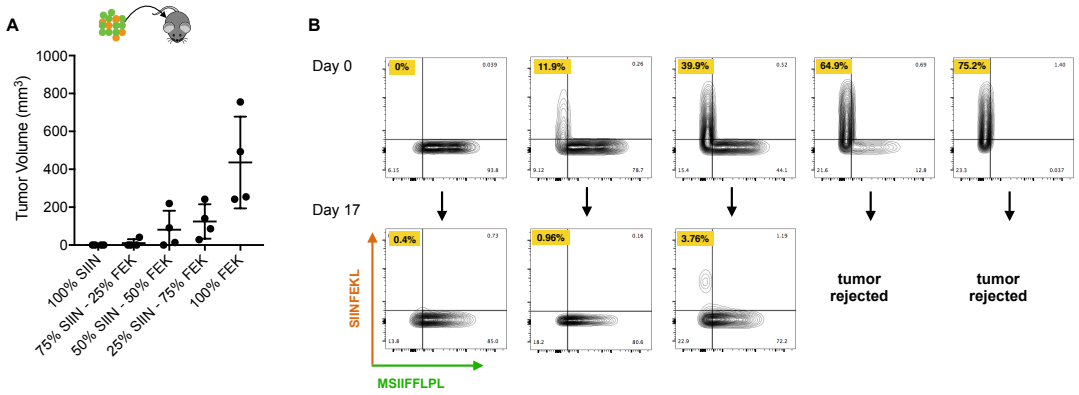


Figure 36 Limited bystander killing of growing tumors in the RMA/S antigen minigene tumor model. (A) Mixtures of $5x10^6$ RMA/S PresentER-SIINFEKL and PresentER-FEKIILSN injected into wild type mice. Tumor sizes at day 15 are presented because some animals had to be sacrificed on day 17. **(B)** $CD45.1^+$ mice were injected with 5 mixtures of PresentER-SIINFEKL (mCherry) and PresentER-MSIIFFLPL (GFP) cells at several ratios. The top row shows the percentage of SIINFEKL and MSIIFFLPL cells at time 0. Tumors were harvested at day 17, enzymatically disaggregated and the percentage of $CD45.2^+$ $eGFP^+$ and CD45.2 $mCherry^+$ cells were quantified (bottom row). The percentage of PresentER-SIINFEKL cells in each pretumor and tumor sample is highlighted in yellow. The two tumors with the highest percentage of PresentER-SIINFEKL cells were complete rejected and no cells could be recovered.

Library screen in vivo reveals the limitations of the immune system to eliminate immunogenic tumor subclones

We hypothesized that a pooled screen *in vivo* might reveal the determinants of immunogenic MHC-I ligands if a tumor bearing a library of MHC-I antigen were depleted of cells bearing immunogenic peptide-MHC while tumor cells bearing non-immunogenic pMHC were spared (**Figure 37A)**. Using the PresentER system, such a screen could be done on large scale and identify hundreds or thousands of immunogenic antigens at once, in contrast to identifying immunogenic antigens one-by-one. Immunoediting *in vivo* leading to loss of tumor clones with immunogenic neoantigen-encoding mutations has previously been observed in syngeneic mouse tumor models[82-84], suggesting that this approach might be viable.

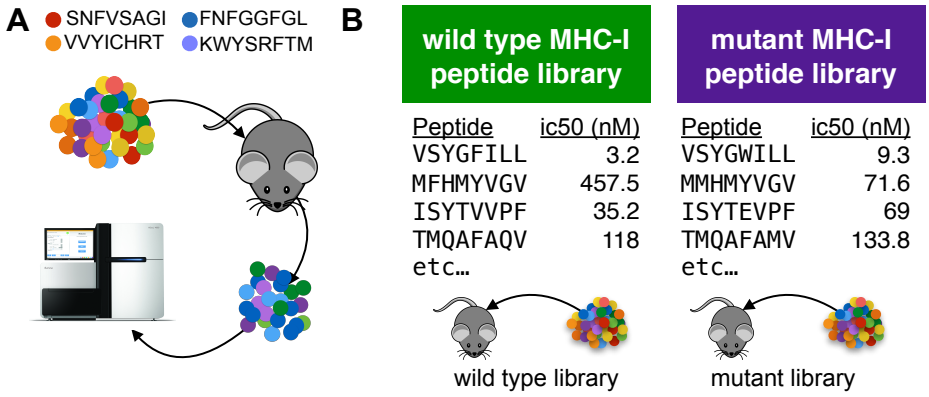


Figure 37 A drop-out screen for MHC-I peptide immunogenicity in wild type mice. Schematic of the drop-out screen for MHC-I immunogenicity. Mice were injected with mixtures of RMA/S cells, each of which expresses a different peptide that also served as a genetic barcode for later deconvolution. **(B)** Two libraries of mouse MHC-I peptides were constructed. Left: Wild type peptides identified by searching the mouse proteome for peptides predicted to bind to MHC-I (NetMHCpan H-2Kb ic50 < 500nM). Right: single amino acid mutants of each of the wild type peptides.

We searched across all mouse proteins and randomly selected 5,000 8-mer peptides that were predicted by NetMHCPan to bind to the B6 mouse MHC-I allele H-2Kb. We

also designed a library of mutated peptides by selecting single amino acid substituents of each peptide in the wild-type peptide library that did not eliminate MHC-I binding (**Figure 37B)**. The MHC-I affinities and properties of the mutated peptides are further described in **Figure 38**. On average, the mutated libraries have slightly higher affinity for MHC-I than the wild-type peptides but the plurality of mutated peptides are within 100nM of their non-mutated counterparts **(Figure 38A-B)**. The residues that are changed from the wild-type to the mutant library tend to be at positions 4, 6 and 7 (**Figure 38C)**. The majority of mutations are isomorphic (polar>polar, hydrophobic>hydrophobic and charged>charged), but ~⅓ of peptides feature a hydrophobic>other substitution (**Figure 38D)**. Several known immunogenic and known non-immunogenic control antigen minigenes were included in each library, including some that were described in Figure 1. The libraries of wild type and mutant libraries were separately cloned and introduced into RMA/S cells by transduction at multiplicity of infection < 0.3, thereby ensuring that few cells received more than one minigene. Naive C57BL6/N mice were injected with either (a) $5x10^6$ cells expressing the wild type library, (b) $5x10^6$ cells expressing the mutant library, (c) $5x10^6$ mutant library cells plus $1x10^6$ wild type RMA/S ("padded" to provide a buffer against bystander killing in the event that most of the cells in the tumor were immunogenic) or (d) $5x10^6$ wild type library cells mixed with $5x10^6$ mutant library cells (n=5 per group). The tumors were allowed to grow for 17 days and then harvested (**Figure 39**). Genomic DNA was extracted from all of the tumors, as well as RMA/S library cells frozen on the day of injection ("pretumor" samples) and the minigenes encoded by each cell were amplified by PCR and sequenced by Illumina next generation sequencing.

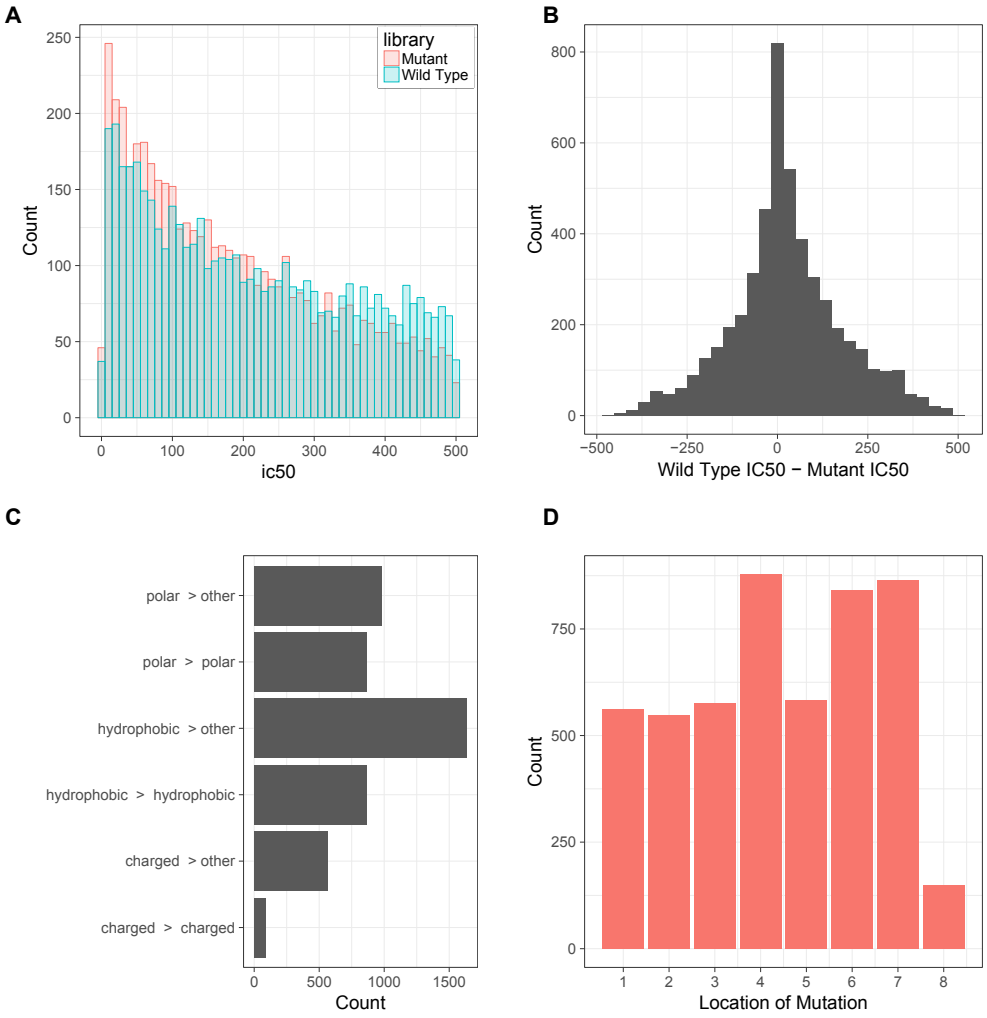


Figure 38 Characteristics of the wild-type and mutated antigen minigene libraries. Characteristics of the wild-type and mutated antigen minigene libraries. **(A)** Distribution of MHC-I affinities in the wild-type and mutated library. **(B)** Difference in NetMHCPan predicted H2-Kb affinity (ic50) between wild-type and mutant peptides **(C)** The type of amino acid changes in the mutant library (Polar: Q,N,H,S,T,Y,C,M,W; Charged: R,K,D,E; Hydrophobic: A,I,L,F,V,P,G). **(D)** The positions of the mutations in the mutant library with respect to the wild-type 8-mer.

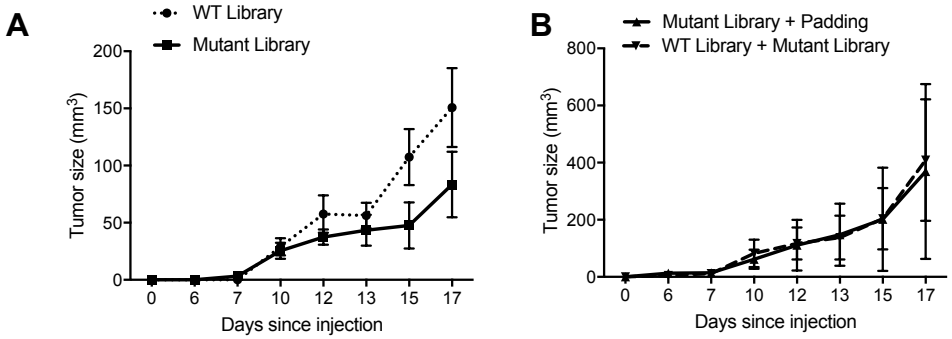


Figure 39 Growth curves of RMA/S tumors bearing libraries of PresentER minigenes. C57BL6/N mice (n=5 per group) were injected with RMA/S cells expressing PresentER libraries: **(A)** $5x10^6$ WT library cells or $5x10^6$ mutant library cells **(B)** $5x10^6$ mutated peptide library RMA/S cells plus $1x10^6$ untransduced RMA/S cells ("padded" mutant peptide library) or 10^7 wild type library and mutated library cells. Tumor growth curves are plotted.

Comparison of minigene abundance in the tumor outgrowth with minigene abundance in the pre-tumor samples led to the surprising finding that no minigenes were robustly depleted during growth *in vivo*. (**Figure 40**). This result stood in stark contrast to the experiments described above, in which mutant peptide expressing clones were efficiently depleted in immunocompetent mice. Some minigenes that were not abundant in the library (<1/10,000) at the time of injection were depleted or dropped out entirely in the tumor due simply to stochastic drop-out; however, minigenes that were abundant in the library at time of injection maintained their abundance despite *in vivo* growth of the tumor. Surprisingly, even positive control minigenes encoding strongly immunogenic peptides (orange points) were not depleted in this context. Analysis of minigene abundance in individual animal tumors (as opposed to the average of several tumors) yielded the same conclusion that few, if any, minigenes were reliably depleted (**Figure 41**).

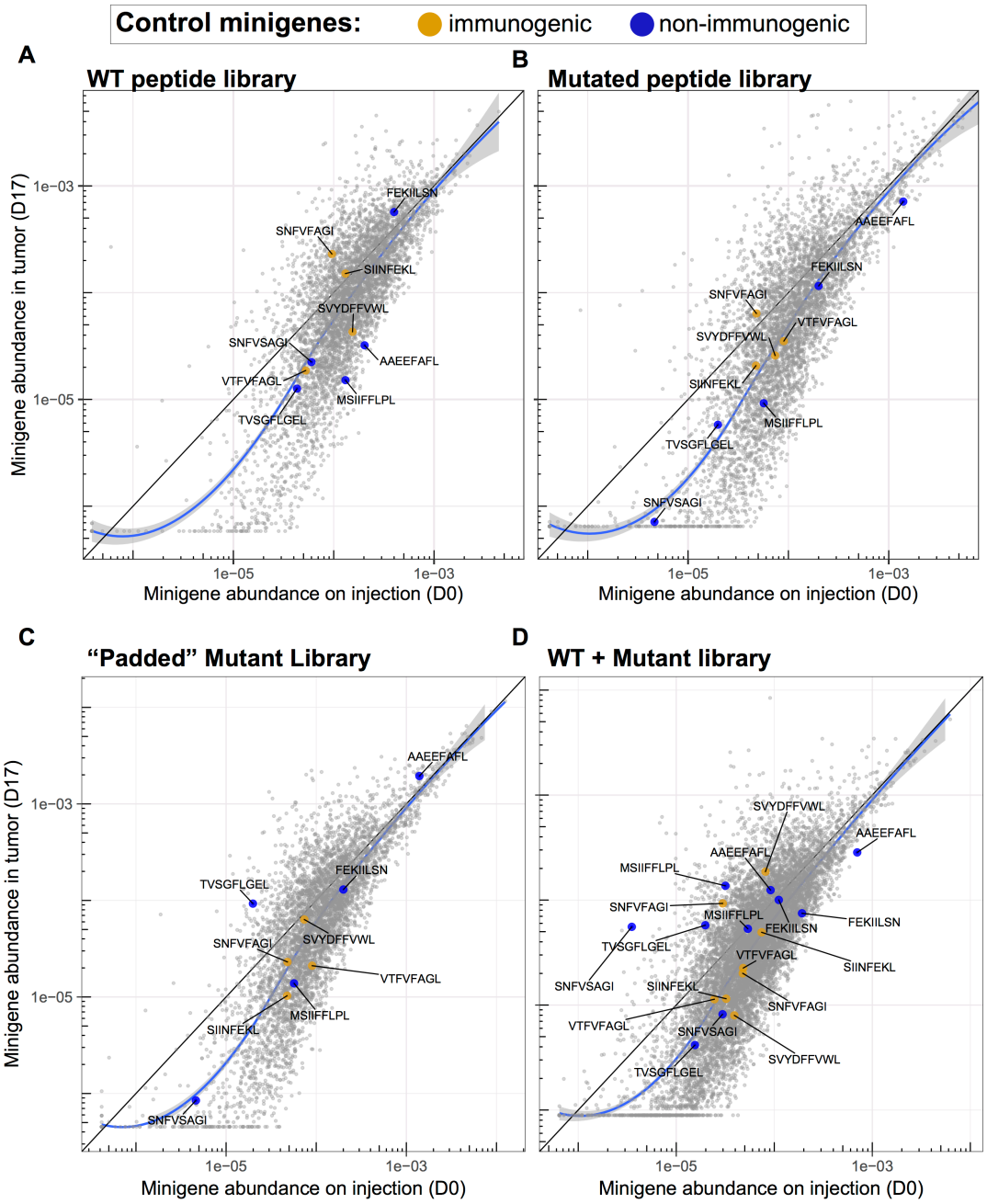
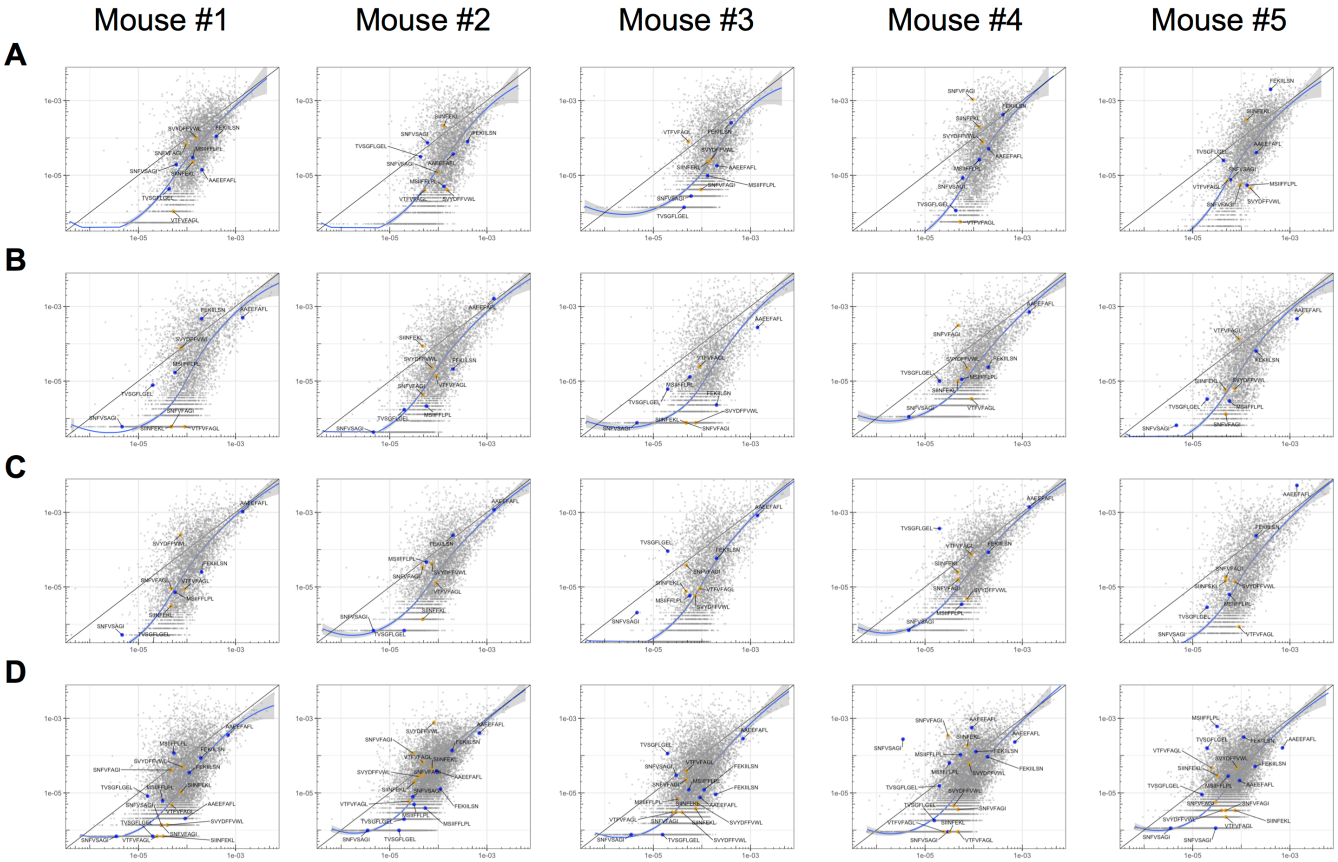


Figure 40 Abundance of each minigene in RMA/S tumors after growth *in vivo.* A scatter plot showing the average frequency of each minigene in the library before injection of the cells (x-axis) and after growth of the tumor in wild type mice (y-axis). The abundance of each minigene before injection (n=3) is plotted on the x-axis while the abundance of each minigene after 17 days of growth in a wild type mouse is plotted on the y-axis (n=5). Orange circles indicate positive control (immunogenic) minigenes; blue circles indicate negative control (non-immunogenic) minigenes. The straight black lines indicate x=y. LOESS (local best fit) lines are plotted in blue.

Figure 41 Abundance of each minigene in each mouse tumor after growth *in vivo.* The abundance of each minigene in each mouse tumor after 17 days of growth in vivo (y-axis) compared to the abundance of each minigene in culture (x-axis) before injection across four groups of tumors: **(A)** wild type, **(B)** mutated, **(C)** "padded" mutated and **(D)** mixed wild type and mutated peptide libraries. The straight black lines indicate x=y. LOESS (local best fit) lines are plotted in blue.


Mouse #1
Mouse #2
Mouse #3
Mouse #4
Mouse #5
A
B
C
D

The absolute number of cells at time of injection can be estimated from the relative abundance of each minigene in the pre-tumor samples. The least abundant minigenes were found at 1 per $5x10^6$ cells (0.00002%) and the most abundant at ~$23x10^3$ per $5x10^6$ (~0.5%). The immunogenic controls SIINFEKL, VTFVFAGL and SNFVFAGI were found at 0.005% - 0.013% of cells before injection into mice, which correspond to between 250 and 650 cells injected out of $5x10^6$. The number of cells expressing the immunogenic antigens upon injection is very low and thus analogous to the number of cells that are present during early tumor development, when immunogenic transformed cells first arise. The surprising inability of the immune system to eliminate these demonstrably immunogenic cells may shed light on the limits of immune cell activation and killing during the early stages of tumorigenesis that enable outgrowth and escape of immunogenic cancers.

Vaccination with minigene library does not result in immunosurveillance

A possible explanation for failure of immune surveillance in our RMA/S MHC-I minigene library model might be insufficient antigen available in the growing tumor to activate an initially productive anti-tumor immune response in naive mice. If this were true, we hypothesized that T cells from antigen experienced mice might be able to detect and kill immunogenic cells in a highly heterogeneous tumor. Although vaccination with soluble peptides has been shown to generate robust T cell immunity[86,87], this is not cost effective at scale. Alternatively, there is precedence for the idea that vaccination with a library of mutated antigens can lead to immunogenic T cell responses that lead to slower tumor growth or clearance[88]. We decided to vaccinate mice with irradiated tumor cells bearing the library of MHC-I peptides. In

order to avoid confounding immunity to the RMA/S cells themselves[89], we used CRISPR/Cas9 to generate a B6-syngeic *Tap2*$^{-/-}$ MCA205 fibrosarcoma cell line. A single cell clone of *Tap2* knockout MCA205 was selected and *Tap2* knockout was validated by RT-PCR and next generation sequencing. Decreased surface MHC-I staining was expected and observed, because the TAP complex is a key chaperone of peptide/MHC-I complex formation (**Figure 42**).

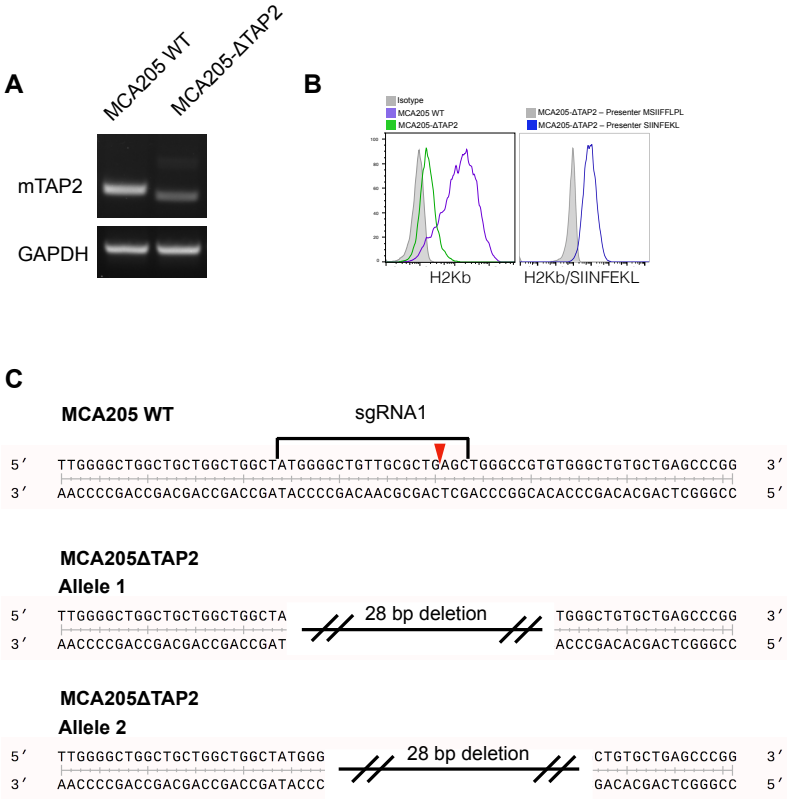


Figure 42 Generation of an MCA205 TAP deficient cell line.
MCA205ΔTap2 cell line was generated by transient transfection of MCA205 cells with a plasmid encoding Cas9 and an sgRNA directed at Tap2. A single cell clone with an INDEL in both alleles of Tap2 was selected and expanded. **(A)** RT-PCR of genomic DNA shows a smaller band than wild-type cells. **(B)** Left: Loss of surface H2-Kb in the ΔTap2 line. Right: transduction of MCA205ΔTAP2 with PresentER-SIINFEKL or PresentER-MSIIFFLPL and staining with SIINFEKL/H2-Kb TCR mimic antibody shows ER independent presentation of the SIINFEKL antigen. **(C)** Next generation sequence of the MCA205ΔTap2 single cell clone shows two different 28bp deletions in both Tap2 loci.

WT B6 mice were vaccinated three times, once every 5 days, with $1x10^7$ irradiated MCA205Δ*Tap2* cells bearing the wild-type library minigenes (**Figure 43A).** Splenocytes and draining lymph nodes from three vaccinated and three non-vaccinated mice were examined at day 17 after the final vaccination and analyzed for the presence of antigen experienced T cells. Five control peptide tetramers were used, three of which are immunogenic and were present in the library (SIINFEKL, SNFVFAGI, VTFVFAGL), one which is not immunogenic but was present in the library (MSIIFFLPL) and one which is immunogenic but not found in the library (SIYRYYGL). Only the immunogenic peptides found in the library showed an increased number of $CD44^+$/$tetramer^+$ CD8 T cells, while the other two peptides did not show significant changes (**Figure 43B**). Therefore, vaccination with the library yielded detectable specific T cell populations to the immunogenic peptides.

Vaccinated and non-vaccinated mice were then challenged with $5x10^6$ RMA/S cells bearing the wild type peptide library. Slower tumor growth was noted in the vaccinated as compared to the non-vaccinated mice, suggesting that a vaccine-related anti-tumor effect may have occurred (**Figure 43C**). However, neither vaccinated nor unvaccinated animals showed depletion of immunogenic control minigenes in relation to the non-immunogenic control minigenes (**Figure 44**). Thus, although slower tumor growth was noted, antigen-specific immunity was not observed in response to prophylactic vaccination in the library setting, suggesting a possible response to some other, broadly expressed, cellular antigens not represented in the peptide library.

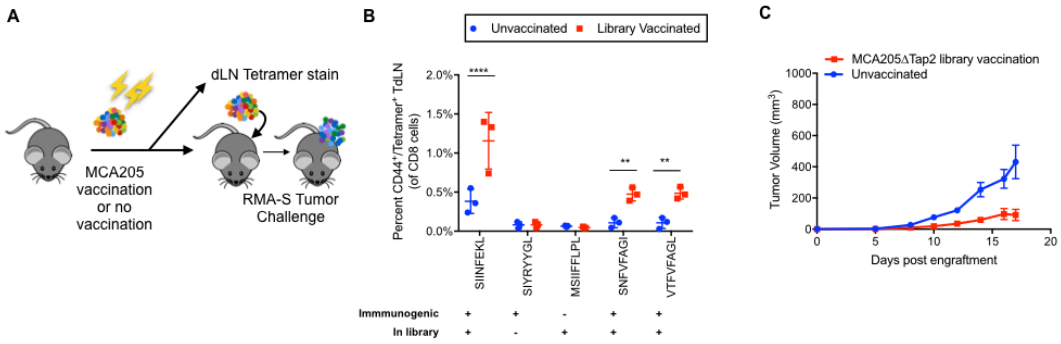
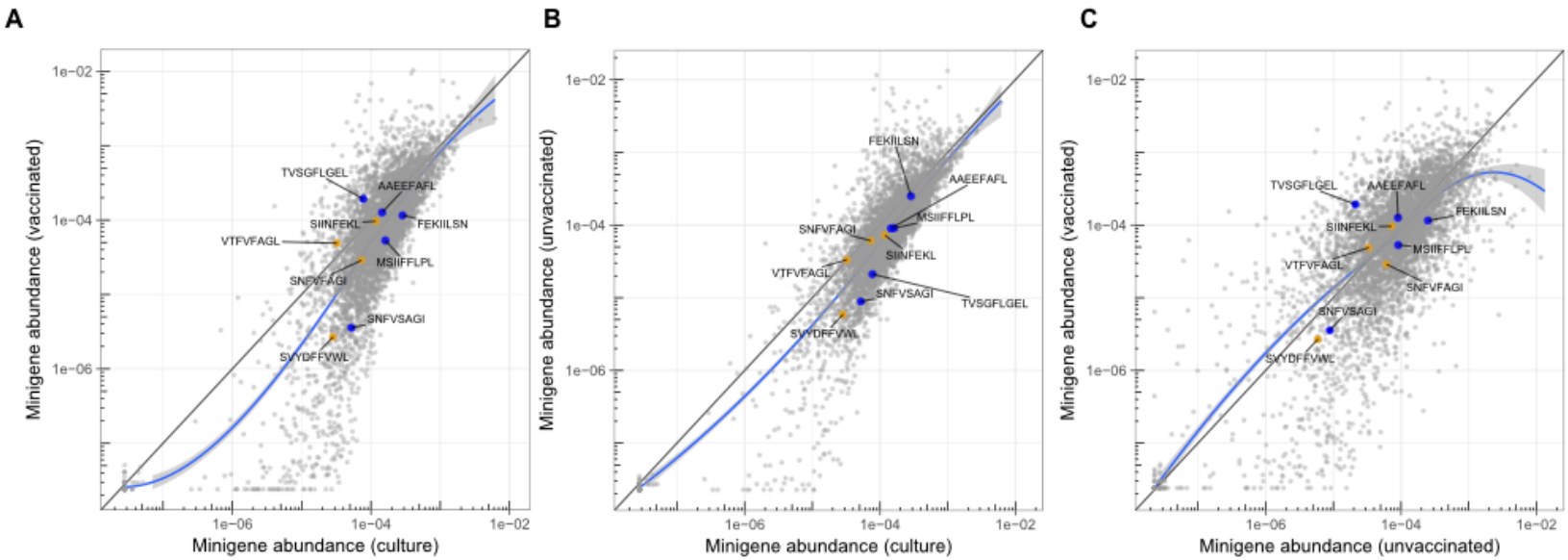


Figure 43 Vaccination of wild type mice with minigene library-expressing MCA205ΔTap2 cells leads to increased antigen-reactive T cells. Schematic of the vaccinations performed on C57BL/6N mice. 10^7 Irradiated MCA205ΔTap2 cells expressing wild type library peptides were injected subcutaneously every six days (for a total of three vaccinations). On day 18, three mice from each group were sacrificed for tetramer analysis. Draining lymph nodes and splenocytes were stained with H-2Kb peptide tetramers. At day 18, five mice were challenged with $5x10^6$ RMA-S cells expressing the library. **(B)** Splenocytes and draining lymph node cells from vaccinated animals were stained for CD8, CD44, and H-2Kb/peptide tetramers. Five control peptides were evaluated: 4 found in the library and 1 peptide not found in the library. The frequency of CD44/tetramer positive CD8 cells is reported. **(C)** Growth curves of RMA/S library tumors in in vaccinated or unvaccinated mice.

Figure 44 Vaccination of wild type mice with minigene library-expressing MCA205ΔTap2 cells does not increased immune surveillance. Average abundance of each minigene (n=5 unvaccinated mice and n=4 vaccinated mice) in cultured cells before injection into mice (x-axis) compared to minigene abundance in tumors harvested from vaccinated (y-axis) **(A)** or non-vaccinated (y-axis) **(B)** mice. Each circle is a minigene. Orange circles indicate positive control (immunogenic) minigenes; blue circles indicate negative control (non-immunogenic) minigenes. **(C)** Direct comparison of minigene abundance in tumors grown in vaccinated and non-vaccinated animals. The straight black lines indicate x=y. LOESS (local best fit) lines are plotted in blue.


A
B
C
Minigene abundance (vaccinated)
Minigene abundance (culture)
Minigene abundance (unvaccinated)
Minigene abundance (culture)
Minigene abundance (vaccinated)
Minigene abundance (unvaccinated)
1e−02
1e−04
1e−06
TVSGFLGEL
AAEEFAFL
SIINFEKL
FEKIILSN
VTFVFAGL
MSIIFFLPL
SNFVFAGI
SNFVSAGI
SVYDFFVWL

Rejection of immunogenic subclones depends on subclone fraction in tumor

While we have demonstrated that PresentER minigenes peptides can generate effective antigen-specific immunogenicity in bulk RMA/S tumor assays, the same response does not occur in tumors bearing libraries of MHC-I ligands. In order to test if there is a threshold level of tumor cell clonality necessary to effectively activate the immune system, CD45.1 mice were injected with mixtures of immunogenic (labeled with eGFP) and non-immunogenic (labeled with mCherry) RMA/S cells and flow cytometry was performed on reisolated tumors 17 days later (). Relative to their proportion upon engraftment, tumor cells bearing the immunogenic peptides SIINFEKL and TAYRYHLL were depleted when they comprised as little as 1% of the tumor. Below 1%, depletion of cells bearing these two minigenes could not be detected. Depletion of cells bearing the SNFVSAGI peptide could be reliably detected at 50% and in some tumors at 10%, however depletion could not be detected when the SNFVSAGI cells were found at less than 10% of the tumor (**Figure 45**). These findings are surprising, as they indicate that immunogenic tumor subclones can persist within a tumor and that rejection or persistence is dependent on tumor cell percentage of the total tumor mass during tumorigenesis, and not immunogenicity of the cell alone.

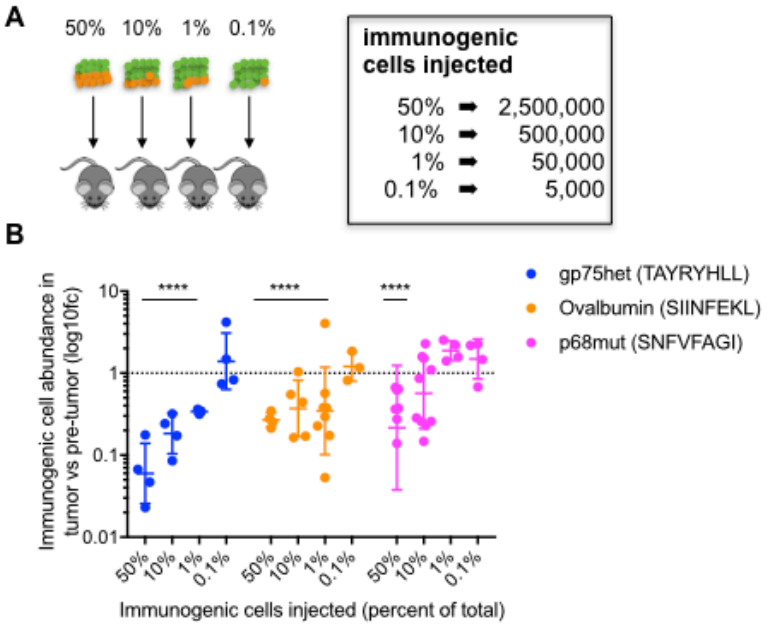
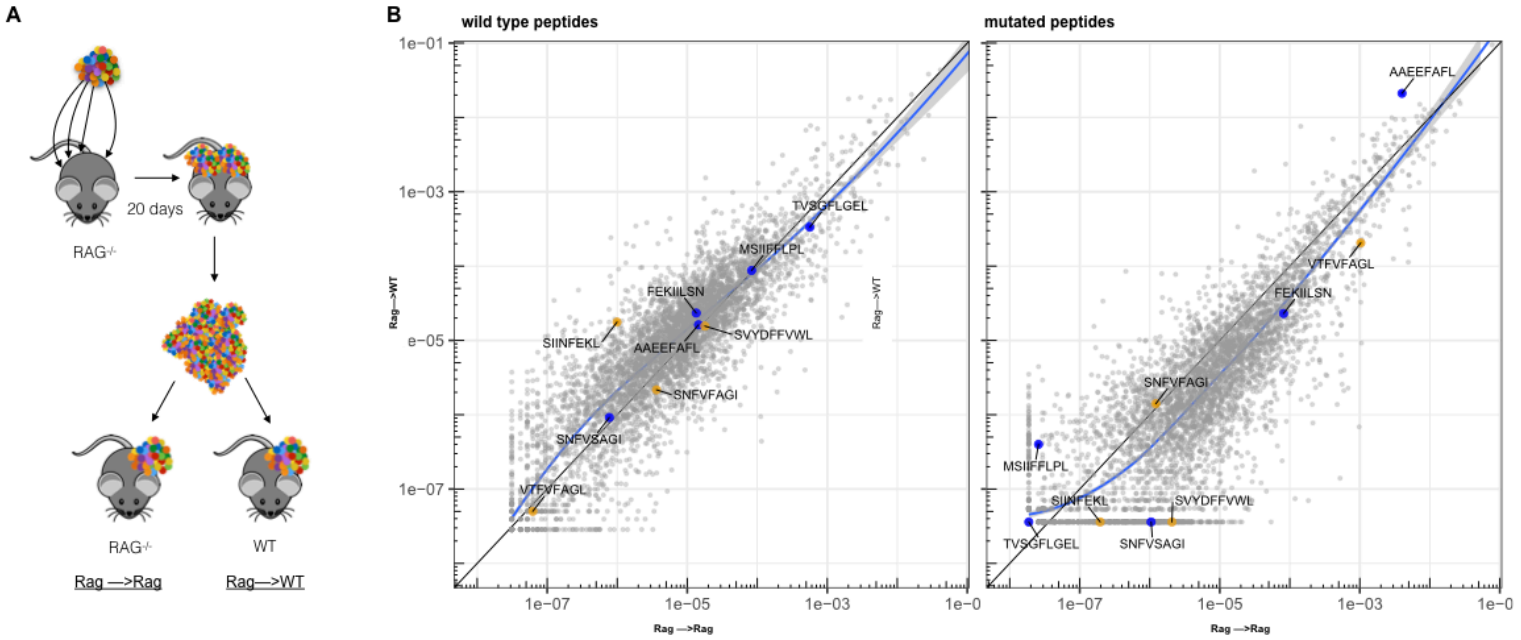


Figure 45 Immunosurveillance fails when tumor subclone frequency is low. (A) Schematic and **(B)** results of experiments to detect antigen-specific immune surveillance thresholds. CD45.1^{+} C57BL/6N mice were injected with mixtures of immunogenic and non-immunogenic RMA/S cells. Mixtures were 50%, 90%, 99% or 99.9% non-immunogenic PresentER-MSIIFFLPL (mCherry) mixed with 50%, 10%, 1% or 0.1% cells expressing one of three different immunogenic minigenes (eGFP) noted in the inset. After 17 days in the mouse, tumors were enzymatically disaggregated, stained with CD45.2 and analyzed by flow cytometry. t-tests with the alternative hypothesis that log10 fold change is 0 were performed for each group. Some tumors were rejected entirely and these were excluded from the presented depletion analysis. At least 3 tumors per group yielded sufficient numbers of cells to be confident that depletion had occurred. All groups marked with 4 stars reached a p-value of <0.001.

Failure of the immune system to eliminate immunogenic subclones present at low fractional abundance could either be due to the low quantity of total antigen present in the tumor or to the low percentage of cells bearing each antigen. In order to discriminate between these two possibilities, we increased the amount of tumor injected, thus increasing the total amount of antigen the immune system sees while keeping the percentage of each antigen within the tumor the same. We grew RMA/S

library tumors in RAG$^{-/-}$ mice, harvested the tumors at 20 days and retained a portion of the material for sequencing. The rest of the tumor material was transferred into the flank of WT B6 or RAG-/- mice. Each animal received approximately 1 milliliter of tumor fragments (~2.5x10^{8} cells), which represents a 40-100 fold increase in cells expressing each antigen (**Figure 46A**). After 17 days, the transferred tumors were harvested and sequenced. Once again, as we observed in both naive and vaccinated mice, robust depletion of immunogenic minigene peptides did not occur. Overall minigene abundance was highly correlated between the tumor material transferred and the tumors harvested 17 days later in both WT and RAG-/- mice and in both the wild type and mutant libraries (**Figure 46B**). These results reveal that it is the percentage of tumor cells bearing each antigen within a tumor and not the total quantity of antigen that determines if an effective immune response occurs.

Figure 46 Large numbers of tumor cells do not overcome failure of immunosurveillance when tumor subclone frequency is low. (A) In order to overcome the hypothesized lack of immune surveillance at low minigene abundance, 2 $RAG^{-/-}$ mice were injected at 4 sites with $5x10^6$ RMA/S library cells per site. Tumors were harvested after 20 days, minced and pooled. Approximately $1mm^3$ of tumor fragments were implanted subcutaneously into the flank of either wild-type or $RAG^{-/-}$ mice and allowed to grow for 17 days. **(B)** The abundance of each minigene (average of 3 mice) in tumors transferred to WT mice (y-axis) compared to tumors transferred to $RAG^{-/-}$ mice (x-axis) is plotted. Each circle is a minigene. Orange circles indicate positive control (immunogenic) minigenes; blue circles indicate negative control (non-immunogenic) minigenes. The straight black lines indicate x=y. LOESS (local best fit) lines are plotted in blue.


A
20 days
RAG-/-
RAG-/-
WT
Rag —>Rag
Rag—>WT
B
wild type peptides
mutated peptides
1e−01
1e−03
e−05
1e−07
Rag—>WT
Rag —>Rag
TVSGFLGEL
MSIIFFLPL
FEKIILSN
SIINFEKL
AAEEFAFL
SVYDFFVWL
SNFVFAGI
SNFVSAGI
VTFVFAGL
1e−07
1e−05
1e−03
1e−0

Discussion

The immune system is capable of recognizing cancer cells as foreign based on the presentation of altered or unusual MHC-I ligands on the surface of tumor cells—a phenomenon that has been leveraged for cancer therapy such as immune checkpoint blockade and adoptive cell transfer[19,21,66]. Despite this recognition, the immune system fails to clear these immunogenic, clinically detected tumors on its own, as has been paradoxically noted for decades[67]. It is not clear at what point in tumorigenesis the immune system begins to initiate (ineffectively) a response to immunogenic MHC-I ligands presented on cancer cells. There is epidemiologic evidence that immunocompromised patients develop more tumors, suggesting that the immune system may prevent some tumors (mostly virally driven) from ever manifesting clinically by effective immunosurveillance. However, recent data also suggests that mutations accumulate at high rates in sun-damaged, but otherwise normal, tissue at levels comparable to cancer cells[90] and that tumors face overall little negative selective pressure[91]. Indeed, immune escape by loss of MHC [92-96], when it occurs, is a late and subclonal event[97,98]. If T cell surveillance were highly effective during early tumor development, negative selective pressure would be notable in the evolutionary trajectories of tumors and HLA loss would be expected to occur early, frequently and in a clonal manner. We interpret the data from human and animal experiments to suggest that routine T cell immunosurveillance of nascent cancer cells does occur, but is limited by unknown factors and that recognition of tumors as foreign occurs mostly later in tumor development.

If tumor immunosurveillance is sometimes effective in the early growth of a tumor, we hypothesized that it might be due to potent neoantigens expressed by cancer cells. Discovery of the biochemical characteristics of mutationally-derived neoantigens that are immunogenic would be important to clarify why some neoantigens are tolerated by the immune system and others are not. The significance of this question is underscored by the major challenge currently facing cancer immunotherapy: the identification of patients whose tumors bear immunogenic neoantigens—and are thus likely to respond well to immune checkpoint blockade or other immunotherapies—vs those patients who will be unnecessarily exposed to potentially toxic therapies without a chance for efficacy.

Here, we have developed a reductionist approach to determine which MHC-I ligands are immunogenic in immunocompetent mice. We demonstrate that tumors are rejected when an immunogenic pMHC is expressed on all or most injected cancer cells. To our surprise, we have discovered that effective T cell responses are not mounted against highly immunogenic peptides when cells expressing these peptides are a minor fraction of the tumor, which is the case in early developing tumors in humans. Furthermore, the threshold percentage of tumor cells necessary to yield an antigen-specific immunogenic T cell response varies with the antigen. We propose that ineffective T cell responses may be a consequence of intratumoral (or intra-tissue) heterogeneity and that neoantigen clonal fraction is an important, overlooked aspect of MHC-I antigen immunogenicity.

There are many mechanisms by which immune mediated killing of immunogenic cancer cells can fail. Here we report the first evidence that tumor heterogeneity is an

explicit factor leading to failure of effective immunity in a tumor and that the maximum level of heterogeneity tolerable before immune escape occurs is dependent on the antigen. Immune checkpoints or active tumor suppression do not explain the observed immune evasion because such mechanisms would be expected to suppress T cell killing irrespective of the fraction of tumor cells that expresses an immunogenic epitope and might be even more effective for minor subclones. This is evidenced by robust depletion of the highly immunogenic peptides SIINFEKL and TAYRYHLL when present at ≥1% of the tumor. The mechanism by which T cell responses are restrained when immunogenic MHC-I antigens are present at low frequency is not yet clear. Low levels of antigen presentation may lead to ineffective cross-presentation of antigen in the tumor draining lymph node, thus limiting T cell activation[99]. Moreover, in animal models, low levels of immunogenic epitopes of oncogenic drivers presented on transformed cells early during tumorigenesis (at the premalignant stage) were shown to induce a program of cellular hypo-responsiveness in tumor-specific (oncogene-specific) CD8 T cells [100,101]. After prophylactic vaccination with irradiated library cells we observed increased numbers of antigen-specific T cells in splenocytes and draining lymph nodes, suggesting that some antigen-specific T cell proliferation does occur, but it does not lead to a productive anti-tumor response. We speculate that the mechanism of immune escape in this model is either ineffective T cell activation or failure of activated T cells to identify and kill antigen-positive cells present at low abundance within a sea of cells displaying irrelevant antigens.

New therapeutic modalities such as immune checkpoint blockade have shown clinical efficacy in the treatment of human tumors, but the toxicity of the drug regimens has led to many efforts to find biomarkers that predict patient response. Tumors with high

mutation burdens—which are more likely to have immunogenic neoantigens[102-104]—and tumors that have mismatch repair (MMR) deficiency or microsatellite instability (MSI-H) respond well to checkpoint blockade[105,106]. In mouse models, syngeneic tumors with MMR gene knock outs accumulate mutations over time and grow more slowly in wild type mice than do tumors without MMR deficiency. Cell lines derived by sub-cloning of MMR deficiency lines—thus increasing the clonal fractions of each neoantigen—grow more slowly or are rejected entirely. In all cases, MMR inactivation in tumors leads to better responses to checkpoint blockade than the parental tumors[107]. Moreover, survival is inversely related to tumor neoantigen clonality in human patients with lung adenocarcinoma[74,75]. Patients whose tumors bear high numbers of subclonal (or branched) neoantigens have increased levels of relapse and worse survival than those patients with more homogenous tumors[74,75]. Response to checkpoint blockade in lung and skin tumors is also associated with lower levels of intratumoral heterogeneity[74]. The combination of these data are highly suggestive that while total neoantigen burden is important for long term survival and response to checkpoint blockade, neoantigen clonality is an additional important factor in mediating tumor regression.

The discovery that intratumoral heterogeneity prevents effective T cell responses has implications for the development of therapeutic cancer vaccines, checkpoint blockade, adoptive T cell therapies, and studies of tumor immunogenicity. In addition, our data may provide a new understanding of mechanisms of immune surveillance and its failure that allows growth and evolution of tumors and subclones. For instance, recent data shows that subclones with very low clonal fraction yet distinct intratumoral functions may be important to tumor survival and growth[108]—suggesting

that even low abundance tumor subclones may be important targets for immunotherapy. In general, we may speculate on one mechanism for cancer escape and progression, in which small early cancers generally do not bear immunogenic epitopes derived from their limited number of driver oncogenic proteins and few passenger mutations. Then as the tumors evolve, the potential neoantigens appearing in subclones may not reach a clonal fraction high enough to breach their antigen-specific immunogenicity thresholds, thereby allowing escape of these otherwise immunogenic clones. This model may help to explain why sun-damaged and aged healthy tissue is replete with mutations, sometimes reaching frequencies seen in human cancers[90,109]. Cumulatively, these findings suggest that effective T cell immunity is restrained in the context of healthy tissue and growing tumors and paint a picture of T cell immunogenicity that is poorly captured by existing models of tissue immunosurveillance.

Thesis Summary

PresentER is a method to identify the targets of T cells and TCR like molecules

Here, we describe a new retroviral minigene that leads to the surface expression of a precisely defined, genetically encoded MHC-I ligand. The minigene is capable of driving antigen presentation from only a single inserted copy of the retroviral vector, thus enabling its use in pooled screens. Using a library of >12,000 MHC-I ligands derived from digestion *in silico* of the human proteome, we screened two soluble TCR mimic antibodies against cells expressing this library. We identified hundreds of cross-reactive ligands, some of which were known to be HLA-A*02:01 peptides presented on real tissue. Several of these cross-reactive ligands were not predictable from the initial characterization of the off-targets of these antibodies.

To extend the use of PresentER to membrane bound TCR present on T cells, we first confirmed that T cells bearing antigen-specific TCR could kill cells expressing their target antigen using the PresentER minigene. Then, we generated a library of 5,000 peptides derived from digestion *in silico* of the human proteome and based on the known binding specificities of the A6 and B7 TCR. These TCR were screened against the library, re-confirming 70-80 and 40, respective, single amino acid substituted targets of these TCRs.

Given the large number of TCR-based therapeutics that are in the pre-clinical or early clinical trial phase, there is pressing need for a method that can be used to systematically identify the off-targets of TCR like agents. Several patients treated with these agents have already died from cross-reactivities[31,110] that are potentially identifiable before treatment.

Future Directions

Identification of targets of Tumor Infiltrating Lymphocytes

One long-standing question in the field of cancer immunology is: what are the targets of tumor infiltrating lymphocytes (TILs)? We have demonstrated that the targets of TCR can be discovered using PresentER screening. We propose that the targets of bulk tumor infiltrating lymphocytes could also be discovered by co-culture of these cells with libraries of cells bearing patient-specific neoantigens. Most human cancers have between 1,000 and 30,000 somatic mutations[111]. Assuming 20 peptides can be derived from each mutation and ~5% of these will bind to one or more of the patient's HLA alleles, the number of mutations that must be tested is approximately equal to the number of somatic changes in the patient's cancer. This number is well within the range of peptides that can be tested in a patient-specific manner.

Degenerate libraries of MHC-I ligands

Biased libraries of MHC-I peptides that have been informed by pre-existing information about the targets of T cells are useful for discovery of off-targets. However, the ability to identify the targets of T cells without any pre-existing information about what the target might look like would be a major advance. We propose that relatively small libraries of PresentER encoded MHC-I peptides could be used to identify the key pairs of residues that a cloned TCR reacts to. For instance, the number of 9-mer peptides where the 2nd and 9th amino acid are fixed as Leu1 and Val9 (to ensure HLA-A*02:01 binding) and the 6th or 7th amino acid is fixed as a leucine, and all other positions are varied, is only ~10,000 peptides. Such a small library might be able to identify key residues that a TCR needs to engage with its target, thus enabling the rapid identification of the actual target of the TCR.

PresentER can be used to study the immunogenicity of MHC-I presented antigens *in vivo*, but rejection of immunogenic tumor clones is limited by clonal fraction

What makes one foreign/mutated MHC-I ligand immunogenic and another tolerated? This is a fundamental question in T cell immunology that we thought we might be able to answer using the PresentER antigen minigene system. We first showed that previously known immunogenic MHC-I peptides could mediate tumor rejection in wild type, immunocompetent mice while non-immunogenic peptides were tolerated and allowed tumor outgrowth. We next showed that the immune rejection was specific to cells within the tumor that expressed the immunogenic peptide. Finally, we cloned libraries of WT and mutated mouse MHC-I ligand libraries and grew tumors bearing thousands of distinct peptides in mice, only to find no robust elimination of even the highly immunogenic control peptides found in the tumor libraries. By making carefully defined heterogeneous tumors of comprised of 0.1-50% immunogenic cells, we were able to show that the fraction of cells expressing an immunogenic antigen must be greater than an antigen-specific threshold in order to see immune-mediated elimination. This unexpected finding has implications for the consideration of neoantigen heterogeneity in human cancers during tumorogenesis and is consistent with pre-existing data showing that clinical outcomes are worse for patients with highly heterogeneous tumors compared to those with homogenous tumors. Failure of the immune system to control immunogenic tumor subclones may be a part of the explanation for the poor outcomes in these patients.

We were unable to identify which MHC-I peptides are immunogenic when growing subcutaneous tumors with thousands of distinct MHC-I ligands within WT mice because the clonal fraction needed for depletion is, at minimum, 1% (in the two most immunogenic antigens we studied). Even relatively small libraries of 100 peptides might fail to identify immunogenic peptides based on the thresholds we have examined. As a result, we believe that at least in the context of solid, subcutaneous tumors, the ability to identify immunogenic peptides will remain limited. Introduction of tumors into mice by other methods, such by IV injection led to death of treated mice within days (data not shown). However, we have not yet tried placing the tumor cells in the IP cavity of a mouse, which might lead to the growth of many small patches of tumor cells, some of which may be well recognized by antigen-specific T cells.

Alternatively, now that we have demonstrated the ability of libraries of cells to be depleted by T cells *in vitro*, it might be possible to identify immunogenic cells by stimulating antigen-specific T cells *in vitro* (perhaps using libraries expressed by MCA205ΔTap2 cells) and then co-culturing these cells with RMA/S libraries (or vice versa). These kinds of approaches might allow the systematic identification of immunogenic neoantigens, a goal of critical importance in the field of T cell immunology.

References

1. Scott, A. M., Wolchok, J. D. & Old, L. J. Antibody therapy of cancer. *Nat. Rev. Cancer* **12,** 278–287 (2012).
2. Townsend, A. & Bodmer, H. Antigen recognition by class I-restricted T lymphocytes. *Annu. Rev. Immunol.* **7,** 601–624 (1989).
3. Hunt, D. F. *et al.* Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* **255,** 1261–1263 (1992).
4. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. & Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Molecular & Cellular Proteomics* **14,** 658–673 (2015).
5. Neefjes, J., Jongsma, M. L. M., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat Rev Immunol* **11,** 823–836 (2011).
6. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res* **28,** 235–242 (2000).
7. Garboczi, D. N. *et al.* Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* **384,** 134–141 (1996).
8. Robinson, J. *et al.* The IMGT/HLA database. *Nucleic Acids Res* **41,** D1222–7 (2013).
9. Paul, S. *et al.* HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. *J Immunol* **191,** 5831–5839 (2013).
10. Murray, N. & McMichael, A. Antigen presentation in virus infection. *Curr Opin Immunol* **4,** 401–407 (1992).

11. Abelin, J. G. *et al.* Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **46,** 315–326 (2017).

12. Hermann, C. *et al.* TAPBPR alters MHC class I peptide presentation by functioning as a peptide exchange catalyst. *Elife* **4,** 26 (2015).

13. Pierini, F. & Lenz, T. L. Divergent allele advantage at human MHC genes: signatures of past and ongoing selection. *Molecular Biology and Evolution* (2018). doi:10.1093/molbev/msy116

14. Mason, D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol. Today* **19,** 395–404 (1998).

15. Sewell, A. K. Why must T cells be cross-reactive? *Nat Rev Immunol* **12,** 669–677 (2012).

16. Singh, N. K. *et al.* Emerging Concepts in TCR Specificity: Rationalizing and (Maybe) Predicting Outcomes. *J Immunol* **199,** 2203–2213 (2017).

17. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat Rev Immunol* **14,** 377–391 (2014).

18. Takaba, H. & Takayanagi, H. The Mechanisms of T Cell Selection in the Thymus. *Trends Immunol.* **38,** 805–816 (2017).

19. Rosenberg, S. A. & Restifo, N. P. Adoptive cell transfer as personalized immunotherapy for human cancer. *Science* **348,** 62–68 (2015).

20. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344,** 641–645 (2014).

21. Zacharakis, N. *et al.* Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat. Med.* **24,** 1–14 (2018).

22. Chapuis, A. G. *et al.* Transferred WT1-reactive CD8+ T cells can mediate antileukemic activity and persist in post-transplant patients. *Sci Transl Med* **5,** 174ra27–174ra27 (2013).

23. Rapoport, A. P. *et al.* NY-ESO-1-specific TCR-engineered T cells mediate sustained antigen-specific antitumor effects in myeloma. *Nat. Med.* **21,** 914–921 (2015).

24. Orlando, D. *et al.* Adoptive Immunotherapy Using PRAME-Specific T Cells in Medulloblastoma. *Cancer Research* **78,** 3337–3349 (2018).

25. Chang, A. Y. *et al.* Opportunities and challenges for TCR mimic antibodies in cancer therapy. *Expert Opin Biol Ther* **16,** 979–987 (2016).

26. Chang, A. Y. *et al.* A therapeutic T cell receptor mimic antibody targets tumor-associated PRAME peptide/HLA-I antigens. *J Clin Invest* **127,** 2705–2718 (2017).

27. Dao, T., Liu, C. & Scheinberg, D. A. Approaching untargetable tumor-associated antigens with antibodies. *Oncoimmunology* **2,** e24678 (2013).

28. Sergeeva, A. *et al.* An anti-PR1/HLA-A2 T-cell receptor-like antibody mediates complement-dependent cytotoxicity against acute myeloid leukemia progenitor cells. *Blood* **117,** 4262–4272 (2011).

29. Jaigirdar, A., Rosenberg, S. A. & Parkhurst, M. A High-avidity WT1-reactive T-Cell Receptor Mediates Recognition of Peptide and Processed Antigen but not Naturally Occurring WT1-positive Tumor Cells. *J. Immunother.* **39,** 105–116 (2016).

30. Dao, T. *et al.* Targeting the intracellular WT1 oncogene product with a therapeutic human antibody. *Sci Transl Med* **5,** 176ra33 (2013).

31. Cameron, B. J. *et al.* Identification of a Titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells.

Sci Transl Med **5,** 197ra103–197ra103 (2013).

32. Kotturi, M. F. *et al.* Of mice and humans: how good are HLA transgenic mice as a model of human immune responses? *Immunome Res* **5,** 3 (2009).
33. Birnbaum, M. E. *et al.* Deconstructing the peptide-MHC specificity of T cell recognition. *Cell* **157,** 1073–1087 (2014).
34. Gee, M. H. *et al.* Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes. *Cell* (2017). doi:10.1016/j.cell.2017.11.043
35. Crawford, F. *et al.* Use of baculovirus MHC/peptide display libraries to characterize T-cell receptor ligands. *Immunol. Rev.* **210,** 156–170 (2006).
36. Hadrup, S. R. *et al.* Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. *Nat Meth* **6,** 520–526 (2009).
37. Bentzen, A. K. *et al.* Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes. *Nat Biotechnol* **34,** 1037–1045 (2016).
38. Adams, J. J. *et al.* Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity. *Nat Immunol* **17,** 87–94 (2016).
39. Nielsen, M., Lundegaard, C., Lund, O. & Kesmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57,** 33–41 (2005).
40. Fortier, M.-H. *et al.* The MHC class I peptide repertoire is molded by the transcriptome. *J Exp Med* **205,** 595–610 (2008).
41. Bassani-Sternberg, M. *et al.* Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery

regulating HLA specificity. *PLoS Comput Biol* **13,** e1005725 (2017).

42. Bassani-Sternberg, M. *et al.* Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* **7,** 13404 (2016).

43. Mommen, G. P. M. *et al.* Expanding the detectable HLA peptide repertoire using electron-transfer/higher-energy collision dissociation (EThcD). *Proc Natl Acad Sci USA* (2014). doi:10.1073/pnas.1321458111

44. Townsend, A. *et al.* Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature* **340,** 443–448 (1989).

45. Bacik, I., Cox, J. H., Anderson, R., Yewdell, J. W. & Bennink, J. R. TAP (transporter associated with antigen processing)-independent presentation of endogenously synthesized peptides is enhanced by endoplasmic reticulum insertion sequences located at the amino- but not carboxyl-terminus of the peptide. *The Journal of Immunology* **152,** 381–387 (1994).

46. Müllbacher, A. *et al.* High peptide affinity for MHC class I does not correlate with immunodominance. *Scand. J. Immunol.* **50,** 420–426 (1999).

47. Hirano, N. *et al.* Efficient presentation of naturally processed HLA class I peptides by artificial antigen-presenting cells for the generation of effective antitumor responses. *Clin. Cancer Res.* **12,** 2967–2975 (2006).

48. Cheever, M. A. *et al.* The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. *Clin. Cancer Res.* **15,** 5323–5337 (2009).

49. Ataie, N. *et al.* Structure of a TCR-Mimic Antibody with Target Predicts Pharmacogenetics. *J. Mol. Biol.* **428,** 194–205 (2016).

50. Borbulevych, O. Y. *et al.* T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. *Immunity* **31,** 885–896

(2009).

51. Davis-Harrison, R. L., Armstrong, K. M. & Baker, B. M. Two different T cell receptors use different thermodynamic strategies to recognize the same peptide/MHC ligand. *J. Mol. Biol.* **346,** 533–550 (2005).

52. Stevanović, S. *et al.* Landscape of immunogenic tumor antigens in successful immunotherapy of virally induced epithelial cancer. *Science* **356,** 200–205 (2017).

53. Viola, A. & Lanzavecchia, A. T cell activation determined by T cell receptor number and tunable thresholds. *Science* **273,** 104–106 (1996).

54. Valitutti, S. & Lanzavecchia, A. Serial triggering of TCRs: a basis for the sensitivity and specificity of antigen recognition. *Immunol. Today* **18,** 299–304 (1997).

55. Fellmann, C. *et al.* An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep* **5,** 1704–1713 (2013).

56. Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61,** 1–13 (2009).

57. Schuster, H. *et al.* The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl Acad Sci USA* **114,** E9942–E9951 (2017).

58. Li, Z. *et al.* Identification of a WT1 protein-derived peptide, WT1, as a HLA-A 0206-restricted, WT1-specific CTL epitope. *Microbiol. Immunol.* **52,** 551–558 (2008).

59. Levin, M. C., Lee, S. M., Morcos, Y., Brady, J. & Stuart, J. Cross-reactivity between immunodominant human T lymphotropic virus type I tax and neurons: implications for molecular mimicry. *J. Infect. Dis.* **186,** 1514–1517 (2002).

60. Hausmann, S. *et al.* Peptide recognition by two HLA-A2/Tax11-19-specific T

cell clones in relationship to their MHC/peptide/TCR crystal structures. *J Immunol* **162,** 5389–5397 (1999).

61. Piepenbrink, K. H., Blevins, S. J., Scott, D. R. & Baker, B. M. The basis for limited specificity and MHC restriction in a T cell receptor interface. *Nat Commun* **4,** 1948 (2013).

62. Borbulevych, O. Y., Piepenbrink, K. H. & Baker, B. M. Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics. *J Immunol* **186,** 2950–2958 (2011).

63. Senra, J. *et al.* Abstract 2562: Affinity-enhanced T-cell receptor (TCR) for adoptive T-cell therapy targeting MAGE-A4. *Cancer Res* **78,** 2562 (2018).

64. Border, E. *et al.* Abstract 2564: Selection of affinity-enhanced T-cell receptors for adoptive T-cell therapy targeting MAGE-A10. *Cancer Res* **78,** 2564 (2018).

65. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ* **47,** 20–33 (2016).

66. Tran, E. *et al.* Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344,** 641–645 (2014).

67. Hellström, I., Hellström, K. E., Pierce, G. E. & Yang, J. P. Cellular and humoral immunity to different types of human neoplasms. *Nature* **220,** 1352–1354 (1968).

68. Heike, M., Blachere, N. E. & Srivastava, P. K. Protective cellular immunity against a spontaneous mammary carcinoma from ras transgenic mice. *Immunobiology* **190,** 411–423 (1994).

69. Dubey, P. *et al.* The immunodominant antigen of an ultraviolet-induced regressor tumor is generated by a somatic point mutation in the DEAD box

helicase p68. *J Exp Med* **185,** 695–705 (1997).

70. Shankaran, V. *et al.* IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **410,** 1107–1111 (2001).

71. Schulz, T. F. Cancer and viral infections in immunocompromised individuals. *Int. J. Cancer* **125,** 1755–1763 (2009).

72. Calis, J. J. A. *et al.* Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Comput Biol* **9,** e1003266 (2013).

73. Duan, F. *et al.* Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med* jem.20141308 (2014). doi:10.1084/jem.20141308

74. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351,** 1463–1469 (2016).

75. Reuben, A. *et al.* TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. *Cancer Discov* **7,** 1088–1097 (2017).

76. Turajlic, S. *et al.* Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell* **173,** 595–610.e11 (2018).

77. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Meth* **11,** 783–784 (2014).

78. Gejman, R. S. *et al.* Prospective identification of cross-reactive human peptide-MHC ligands for T cell receptor based therapies. *bioRxiv* 267047 (2018). doi:10.1101/267047

79. Dyall, R. *et al.* Heteroclitic immunization induces tumor immunity. *J Exp Med* **188,** 1553–1561 (1998).

80. Udaka, K., Wiesmüller, K. H., Kienle, S., Jung, G. & Walden, P. Self-MHC-

restricted peptides recognized by an alloreactive T lymphocyte clone. *The Journal of Immunology* **157,** 670–678 (1996).

81. Wang, M. *et al.* Identification of MHC class I H-2 Kb/Db-restricted immunogenic peptides derived from retinal proteins. *Invest. Ophthalmol. Vis. Sci.* **47,** 3939–3945 (2006).

82. Holler, P. D., Chlewicki, L. K. & Kranz, D. M. TCRs with high affinity for foreign pMHC show self-reactivity. *Nat Immunol* **4,** 55–62 (2003).

83. Matsushita, H. *et al.* Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* **482,** 400–404 (2012).

84. DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F. & Jacks, T. Expression of tumour-specific antigens underlies cancer immunoediting. *Nature* **482,** 405–409 (2012).

85. DuPage, M. *et al.* Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. *Cancer Cell* **19,** 72–85 (2011).

86. Ott, P. A. *et al.* An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547,** 217–221 (2017).

87. Sahin, U. *et al.* Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547,** 222–226 (2017).

88. Engelhorn, M. E. *et al.* Autoimmunity and tumor immunity induced by immune responses to mutations in self. *Nat. Med.* **12,** 198–206 (2006).

89. van Hall, T. *et al.* Selective cytotoxic T-lymphocyte targeting of tumor immune escape variants. *Nat. Med.* **12,** 417–424 (2006).

90. Martincorena, I. *et al.* High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* **348,** 880–886 (2015).

91. Martincorena, I. *et al.* Universal Patterns of Selection in Cancer and Somatic

Tissues. *Cell* **171,** 1029–1041.e21 (2017).

92. Masuda, K. *et al.* Loss or down-regulation of HLA class I expression at the allelic level in freshly isolated leukemic blasts. *Cancer Sci.* **98,** 102–108 (2007).

93. Cabrera, T. *et al.* High frequency of altered HLA class I phenotypes in laryngeal carcinomas. *Hum. Immunol.* **61,** 499–506 (2000).

94. Gudmundsdóttir, I. *et al.* Altered expression of HLA class I antigens in breast cancer: association with prognosis. *Int. J. Cancer* **89,** 500–505 (2000).

95. So, T. *et al.* Haplotype loss of HLA class I antigen as an escape mechanism from immune attack in lung cancer. *Cancer Research* **65,** 5945–5952 (2005).

96. Shukla, S. A. *et al.* Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat Biotechnol* **33,** 1152–1158 (2015).

97. McGranahan, N. *et al.* Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **171,** 1259–1271.e11 (2017).

98. Turajlic, S. *et al.* Tracking Cancer Evolution Reveals Constrained Routes to Metastases: TRACERx Renal. *Cell* **173,** 581–594.e12 (2018).

99. Spiotto, M. T. *et al.* Increasing tumor antigen expression overcomes 'ignorance' to solid tumors via crosspresentation by bone marrow-derived stromal cells. *Immunity* **17,** 737–747 (2002).

100. Willimsky, G. & Blankenstein, T. Sporadic immunogenic tumours avoid destruction by inducing T-cell tolerance. *Nature* **437,** 141–146 (2005).

101. Schietinger, A. *et al.* Tumor-Specific T Cell Dysfunction Is a Dynamic Antigen-Driven Differentiation Program Initiated Early during Tumorigenesis. *Immunity* **45,** 389–401 (2016).

102. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350,** 207–211 (2015).

103. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med* **377,** 2500–2501 (2017).

104. Snyder, A. *et al.* Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N Engl J Med* (2014). doi:10.1056/NEJMoa1406498

105. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372,** 2509–2520 (2015).

106. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357,** 409–413 (2017).

107. Germano, G. *et al.* Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature* **552,** 116–120 (2017).

108. Vinci, M. *et al.* Functional diversity and cooperativity between subclonal populations of pediatric glioblastoma and diffuse intrinsic pontine glioma cells. *Nat. Med.* **9,** 400 (2018).

109. Risques, R. A. & Kennedy, S. R. Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLoS Genet* **14,** e1007108 (2018).

110. Morgan, R. A. *et al.* Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. *J. Immunother.* **36,** 133–151 (2013).

111. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500,** 415–421 (2013).